Handbook of Microscopy

Applications in Materials Science, Solid-State Physics and Chemistry

Edited by S. Amelinckx, D. van Dyck, J. van Landuyt, G. van Tendeloo

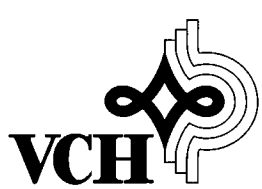
Applications

VCH

S. Amelinckx, D. van Dyck, J. van Landuyt,
G. van Tendeloo

Handbook of Microscopy

Applications



Handbook of Microscopy
Applications in Materials Science, Solid-State Physics and Chemistry

Methods I
1997. ISBN 3-527-29280-2.

Methods II
1997. ISBN 3-527-29473-2.

Applications
1997. ISBN 3-527-29293-4.

Further Reading from VCH

S.N. Magonor, M.-U. Whangbo
Surface Analysis with STM and AFM
Experimental and Theoretical Aspects of Image Analysis
ISBN 3-527-29313-2

D.A. Bonnell
Scanning Tunnelling Microscopy and Spectroscopy
Theory, Techniques and Applications
ISBN 3-527-27920-2

Distribution:
VCH, P.O. Box 10 11 61, D-69451 Weinheim (Federal Republic of Germany)
Switzerland: VCH, P.O. Box, CH-4020 Basel (Switzerland)
United Kingdom and Ireland: VCH (UK) Ltd., 8 Wellington Court, Cambridge CB1 1HZ (England)
USA and Canada: VCH, 333 7th Avenue, New York, NY 10001 (USA)
Japan: VCH, Eikow Building, 10-9 Hongo 1-chome, Bunkyo-ku, Tokyo 113 (Japan)

ISBN 3-527-29293-4

Handbook of Microscopy

Applications in Materials Science, Solid-State Physics and Chemistry

Edited by
S. Amelinckx, D. van Dyck,
J. van Landuyt, G. van Tendeloo

Applications

Weinheim · New York · Basel · Cambridge · Tokyo

Prof. S. Amelinckx
Electron Microscopy for
Materials Science (EMAT)
Universitty of Antwerp - RUCA
Groenenborgerlaan 171
2020 Antwerp
Belgium

Prof. D. van Dyck
Electron Microscopy for
Materials Science (EMAT)
Universitty of Antwerp - RUCA
Groenenborgerlaan 171
2020 Antwerp
Belgium

Prof. J. van Landuyt
Electron Microscopy for
Materials Science (EMAT)
Universitty of Antwerp - RUCA
Groenenborgerlaan 171
2020 Antwerp
Belgium

Prof. G. van Tendeloo
Electron Microscopy for
Materials Science (EMAT)
Universitty of Antwerp - RUCA
Groenenborgerlaan 171
2020 Antwerp
Belgium

Published by
VCH Verlagsgesellschaft mbH, Weinheim (Federal Republic of Germany)

Editorial Directors: Dr. Peter Gregory, Dr. Ute Anton, Dr. Jörn Ritterbusch
Production Manager: Dipl.-Wirt.-Ing. (FH) Hans-Jochen Schmitt

Library of Congress Card No. applied for.

A catalogue record for this book is available from the British Library.

Die Deutsche Bibliothek – CIP-Einheitsaufnahme
Handbook of microscopy : applications in materials science,
solid state physics and chemistry / ed. by S. Amelinckx ... -
Weinheim ; New York ; Basel ; Cambridge ; Tokyo : VCH.
NE: Amelinckx, Severin [Hrsg.]
Applications (1997)
ISBN 3-527-29293-4

Printed on acid-free and chlorine-free paper.

Composition: Alden Bookset, England
Printing: betz-druck, D-64291 Darmstadt
Bookbinding: W. Osswald, D-67433 Neustadt

Short biography of the editors

Severin Amelinckx was born in Willebroek, Belgium, in 1922. He studied at the University of Ghent, receiving his first degree (licence) in mathematics in 1944, his doctorate in physics in 1952, and his aggregation in physics in 1955. Currently he is Emeritus Professor of General Physics and Crystallography associated with the EMAT laboratory of the University of Antwerp (RUCA). Until 1987 he was Director General of the Belgian Nuclear Research Establishment at Mol. He is a member of the European Academy and of the Koninklijke Academie voor Wetenschappen, Letteren en Schone Kunsten van België and former chairman of the division of sciences of this academy.

His research interests include electron diffraction contrast imaging, defects in solids, phase transformations and their resulting domain structures, crystal growth, dislocations, fullerenes and nanotubes, the structure of high-T_c superconductors, modulated structures, and order–disorder in alloys.

Joseph Van Landuyt, who was born in St. Amandsberg, Belgium, in 1938, obtained both his licence (1960) and doctorate in physics (1965) from the University of Ghent. At present he is Professor of General Physics and Crystallography at the University of Antwerp (RUCA and UIA) and of Electron Microscopy at UIA and the Flemish University of Brussels (VUB). He is a member of the Koninklijke Academic voor Wetenschappen, Letteren en Schone Kunsten van België.

His research interests are centered on the study of nanostructural features in alloys, ceramics, and minerals (in particular gems), with special interest in defects in semiconductors and their relation to device performance. More general subjects of interest are structural variants, defects, and phase transitions in various solids.

Gustaaf Van Tendeloo, born in Lier, Belgium, in 1950, received his licence in physics from the University of Brussels (VUB) in 1972, his doctorate from the University of Antwerp (UIA) in 1974, and his aggregation from the University of Brussels (VUB) in 1981. He has been associated with the University of Antwerp (RUCA) since 1972, but has spent extended periods of time as a researcher in Berkeley (USA), Caen (France), and elsewhere. He is currently Professor of Solid-State Physics at the University of Brussels (VUB) and of the Physics of Materials at the University of Antwerp (RUCA and UIA).

His research interests include the electron microscopy of inorganic solids (in particular high-T_c superconductors), solid-state phase transitions, modulated structures, fullerenes, defects in crystals order–disorder in alloys, and nanostructural features in general.

Dirk Van Dyck was born in Wilrijk, Belgium, in 1948. He studied physics, receiving his licence from the University of Brussels (VUB) in 1971 before moving to the University of Antwerp (UIA) for his doctorate (1977) and aggregation (1987). He has been associated with the University of Antwerp since 1971, and is at present Professor of Theoretical Mechanics, Digital Techniques and Image Processing at the University of Antwerp.

Among his research interests are theoretical aspects of dynamic electron diffraction and imaging, holographic reconstruction and structural retrieval, image processing and pattern recognition, and artificial intelligence. In particular, he is involved in the development of a 1 Å resolution microscope in the framework of the Brite/Euram program of the European Union.

The four editors belong to the Electron Microscopy for Materials Science (EMAT) laboratory, University of Antwerp (RUCA), which was founded in 1965. All four have contributed significantly to the development of electron microscopy and its application by numerous publications in journals and books and are members of editorial boards of several international journals in the field of materials science. They have also delivered numerous invited lectures at international conferences in their respective areas of research.

List of Contributors

Barber, David (**I**:2)
Physics Dept.
University of Essex
Colchester
Essex CO4 3SQ
U.K.

Barna, Árpád; Pécz, Béla;
Radnóczi, György (**II**:3)
Research Institute for
Technical Physics of HAS
P.O. Box 76
1325 Budapest
Hungary

Berbezier, Isabelle; Derrien, Jacques (**I**:4)
Centre de Recherches sur les Mécanismes
de la Croissance Cristalline,
CRMC2-CNRS
Campus de Luminy, Case 913
F-13288 Marseille Cedex 9
France

Bernaerts, D.; Amelinckx, Severin (**I**:11)
EMAT, University of Antwerp (RUCA)
Groenenborgerlaan 171
B-2020 Antwerp
Belgium

Blank, H.; Matzke, Hj.; Ray, I. L. F. (**I**:14)
European Commission
European Institute for Transuranium
Elements
Postfach 2340
D-76125 Karlsruhe
Germany

De Hosson, Jeff Th. M.(**I**:1)
Dept. of Applied Physics
Materials Science Centre
University of Groningen
Nijenborgh 4
9747 Groningen
The Netherlands

Ehrlich, K.; Materna-Morris, E.;
Schneider, W. (**I**:14)
Forschungszentrum Karlsruhe
Institut für Materialforschung I
Postfach 3640
D-76021 Karlsruhe
Germany

Gaskell, Philip H. (**I**:9.1)
University of Cambridge
Cavendish Laboratory
Madingley Rd.
Cambridge CB3 0HE
U.K.

Goodhew, Peter J. (**II**:5)
Materials Science and Engineering
University of Liverpool
Liverpool L69 3BX
U.K.

Hubert, A. (**I**:15)
Institut für Werkstoffwissenschaften
Werkstoffe der Elektrotechnik
Universität Erlangen-Nürnberg
Martensstr. 7
D-91058 Erlangen
Germany

Jambers, Wendy;
Van Grieken, René E. (**II**:4)
Micro and Trace Analysis Centre
(MITAC)
Department of Chemistry
University of Antwerp
B-2610 Antwerp
Belgium

Kuo, K. H. (**I**:9.2)
Beijing Lab. of Electron Microscopy
Chinese Academy of Sciences
P.O. Box 2724
100080 Bejing
China

Maußner, H. (**I**:14)
Siemens AG
Power Generating Group (KWU)
Postfach 3220
D-91050 Erlangen
Germany

Oppolzer, H.; Cerva, H.(**I**:3)
Siemens AG
Corporate Research & Development
Otto-Hahn-Ring 6
D-81730 München
Germany

Rühle, Manfred (**I**:6)
MPI für Metallforschung
Seestr. 92
D-70174 Stuttgart
Germany

Salje, Ekhard (**I**:5)
Department of Earth Sciences
and IRC in Superconductivity
Cambridge University
Downing Street
Cambridge CB2 3EQ
U.K.

Schryvers, D. (**I**:1.6)
University of Antwerp (RUCA)
EMAT
Groenenborgerlaan 171
B-2020 Antwerp
Belgium

Van Bockstael, Mark H. G. (**I**:7)
Diamond High Council (HRD)
Institute of Gemmology
Hovenierstraat 22
B-2014 Antwerp
Belgium

Van der Biest, O.; Lust, P; Lambrinou, K.; Ivens, J.; Verpoest, I.; Froyen, L. (**I**:12)
Dept. of Metallurgy and Materials Engineering
Katholieke Universiteit Leuven
De Croylaan 2
B-3001 Heverlee
Belgium

Van Landuyt, Joseph (**I**:7)
University of Antwerp (RUCA)
EMAT
Groenenborgerlaan 171
B-2020 Antwerp
Belgium

Van Royen, Jef (**I**:7)
Diamond High Council (HRD)
Certificates Dept.
Hovenierstraat 22
B-2014 Antwerp
Belgium

Van Tendeloo, Gustaaf (**I**:8 and **I**:1.6)
EMAT
University of Antwerp (RUCA)
Groenenborgerlaan 171
B-2020 Antwerp
Belgium

Voigt-Martin, Ingrid G. (**I**:13)
Institut für Physik
Chemie der Universität Mainz
Jakob-Welder-Weg 11
D-55099 Mainz
Germany

Warlimont, Hans (**II**:2)
Institut für Festkörper- und
Werkstoffforschung e.V
Postfach 27 00 16
D-01171 Dresden
Germany

Yasuda, Katsuhiro; Hisatsune, Kunihiro; Takahashi, Hiroshi; Udoh, Koh-Ichi; Tanaka, Yasuhiro (**I**:10)
Dept. of Dental Materials Science
Nagasaki University
School of Dentistry
1–7-1 Sahamoto
Nagasaki 852
Japan

Zandbergen, H.W.; Træholt, Chresten (**II**:1)
Lab. voor Materiaalkunde
Techn. Universiteit Delft
Rotterdamsweg 137
2628 AL Delft
The Netherlands

Outline

Volume 1: Methods I

Volume 3: Applications

I Classes of Materials

Contents

Volume 3: Applications

II Special Topics

Handbook of Microscopy
S. Amelinckx, D. van Dyck,J. van Landuyt , G. van Tendeloo

General Introduction

The importance of microscopic imaging has in recent years been recognized repeatedly by the awarding of Nobel prizes to the inventors of a number of such methods.

As a consequence of the decreasing scale of many devices, high resolution characterization methods have become of vital importance for further development in these areas. Recent advances in data processing have made it possible to develop imaging modes for a number of methods of chemical analysis, based on particle beams; they have been considered as forms of microscopy, particularly as they are often accessories to microscopic equipment.

The systematic development of new materials strongly relies on their characterization at various and increasing levels of resolution. Structure, microstructure, and defect geometry, as well as chemical composition and spatial distribution are important parameters determining the behavior of materials in practical applications.

At present the materials scientist has a large number of methods at his or her disposal to determine these parameters. In applying these methods, use is made of some kind of probe and the response of the sample to this probe is detected and recorded. In many cases the probe consists of a beam of particles such as neutrons, ions, or electrons, or of electromagnetic radiation such as light, X-rays, microwaves, infrared radiation, or sound waves. However the probe may also be a very fine point or fiber in close proximity to the sample surface leading to some form of interaction (mechanical, optical, electrical, magnetic). The probe may be operated in either a stationary or a scanning mode.

As a guiding principle in selecting the characterization methods to be addressed in the Handbook of Microscopy, we used the requirement that the method should give spatially localized information of the microstructure and/or the composition. Moreover, in order to qualify as 'microscopy', the method should have the potential to provide a magnified real-space image of the sample.

An introductory discussion of the physicochemical principles underlying tthe different methods and the type of information which they can provide is the subject of the first two Volumes of the Handbook, *Methods I* and *Methods II*. The chapters have been written by experienced scientists working in the various fields, the main objective being to provide the reader with sufficient insight and information to allow an optimal choice of the method(s) to be used in order to obtain the desired information.

No materials scientist can master more than a few of these techniques, and presumably would not have access to the

instrumentation necessary to apply a number of them. However, on the basis of these descriptions, it should be possible to judge what results can be expected from the various methods and what their limitations are.

Also problems of image recording and data processing are treated in *Methods II*, in relation to the quantitiative evaluation of microscope images in general. The prospective evolution of recently developed methods and the possibilities of new methods still under development are also briefly discussed.

Different classes of materials and different applications of the same material may require different characterization methods; a single method is usually not applicable to all materials. It is therefore meaningful to illustrate the use of the different methods by a number of case studies classified according to the type of material or to its use. This is the objective of the third volume, *Applications*.

The three volumes are complementary. The *Methods I* and *II* volumes mainly address the operator of the instruments and the scientist who wants to understand and interpret the images, while the *Application* volume will be particularly helpful for materials scientists who has to decide which methods looks most promising for their purposes. Together the three volumes form a state of the art account on the subject, essential for the optimal use and application of microscopy techniques.

As the number of available methods grows, increasingly detailed information can be obtained. However, it also becomes increasingly difficult for a materials scientist to make an adequate choice. This Handbook aims to provide a tool to aid in making such a choice possible and to ease access to the vast literature.

The Editors wish to thank the authors for their efforts to provide, in a limited number of pages, clear overviews that are understandable by the nonspecialist.

We are also indebted to the editorial and production staff of VCH for their efficient help in acquiring the manuscripts of so many authors and for turning them finally into finished, well-presented books.

Special thanks are due to Dr P. Gregory, Dr U. Anton and Dr J. Ritterbusch from the Editorial Office for ensuring excellent communications with VCH.

We are grateful for the confidence of the management of VCH in us and for their continuous support.

The Editors:
S. Amelinckx
D. Van Dyck
J. Van Landuyt
G. Van Tendeloo

Part I

Classes of Materials

Handbook of Microscopy
S. Amelinckx, D. van Dyck, J. van Landuyt, G. van Tendeloo

1 Metals and Alloys

1.1 Imaging Metals and Alloys: Introduction

Mechanical properties of metallic systems are particularly structure-sensitive. The microstructural features in turn are determined by chemical composition and processing. Consequently, most advanced microstructural investigations require a microscope with a resolving power of the order of a nanometer or even better. When the need for such a high-resolution microscope became apparent, it was clear that another type of radiation than ordinary light had to be used to make an image of the specimen. The following considerations concerning the radiation were made: Since the resolving power of an imaging microscope is limited to a value approximately equal to the wavelength of the radiation used, the wavelength should be of the order of 1 nm or smaller. Further, it is essential that the radiation can be manipulated in order to form an image. It does not matter whether this is done by lenses or by mirrors. It is in fact this condition that excludes the use of X rays. Finally, it is a prerequisite that the radiation does not destroy the samples. Furthermore, the transmission coefficient of the sample for the radiation used should be significantly greater than zero. These conditions exclude, for instance, ions and low-energy electrons, respectively. It turned out that electrons, with an energy of the order of 100 keV or more, constitute the most suitable form of radiation. This led to the development of the electron microscope. The main disadvantage of electrons is that they are quite strongly absorbed by matter—including air. This implies that the sample thickness is limited to values considerably smaller than 1 μm and that electron beam interaction with the specimen should take place in vacuo.

With all these possibilities available nowadays, it is usually argued that it is necessary to characterize the microstructure at the *highest* levels of resolution possible in order to understand the behavior of materials and to facilitate the design of either new or improved metallic systems [1–3]. The question is whether this is really true. In other words: does transmission electron microscopy really solve materials problems and what level of resolution is it appropriate to look for? Is the highest resolution the best route to understand properties which are structure-sensitive or not?

This section deals with metallic systems. From the start of transmission electron microscopy (TEM) with the observations of Heydenreich of aluminum foils in 1949 [4] a vast amount of work has been

devoted primarily to metallic systems. As a matter of course it would be simply impossible to cover five decades in one single chapter or even in one book. For that reason we took a different approach. The principal motivation for doing TEM on materials is to image microstructural defects and to derive from these observations their influences on the physical properties. However, one should critically evaluate which defects might be relevant, what kind of information is needed, and, most importantly, on what scale? Following these guidelines we left out all the beautiful electron microscopic work that is focused exclusively on the structural aspect of metals and which does not make a linkage to the physical property relationship as such.

Here we will deal mainly with those defects which contribute to the microstructural features in metals and alloys, that is, those which are not in thermodynamic equilibrium, such as two-dimensional and one-dimensional defects, for example, grain boundaries and dislocations, but also three-dimensional defects such as precipitates and noble gas bubbles. The latter are chosen as an illustration of an area of research in metallic systems to which TEM has contributed a lot. It will be shown that in some cases it is crucial to have information on an atomic level available, but this is not always necessary, and it is sometimes more appropriate to image defects on a micrometer scale instead to correlate the structural information to physical properties.

Mechanical strength is strongly affected by internal interfaces such as grain boundaries. For instance, segregation of alloying elements frequently makes the boundaries particularly suitable paths for brittle cracking at low temperatures. Similarly, segregation of alloying elements has a pronounced influence on fractures occurring during creep by cavitation at boundaries. Both the effect of segregated impurities on cohesion and their fast diffusion along grain boundaries play major roles in the recently discovered brittle fracture at high temperatures. Grain boundary cracking is also the main problem encountered in intermetallic compounds, which would otherwise be attractive structural materials for high-temperature applications.

All the boundary phenomena controlling the above-mentioned material properties occur in the region of the core of boundaries, a few atomic spacings wide. Hence, it is the atomic structure of boundaries in alloys which needs to be understood in order to establish the physical mechanisms of boundary phenomena. Direct observations in boundary structures have become possible recently with the advent of high-resolution electron microscopy and scanning tunneling microscopy. General features of the atomic configurations of grain boundaries have been established on the basis of computer modeling of grain boundary structures which were instrumental in advancing the knowledge of boundary structures [5, 6]. The same arguments are valid for the bonding behavior between dissimilar materials such as metals and ceramic, where it is thought that microscopic information at an atomic level on the interface structure is crucial in understanding and predicting the mechanical properties. In this Section we will present some examples of studies on the structure of grain boundaries, on grain boundary segregation, and on the

structure of interfaces between dissimilar materials where sometimes atomic resolution is necessary to solve a particular materials problem but sometimes information on a micrometer scale is more appropriate.

Besides the planar defects we will also concentrate on line defects, that is, dislocations. With respect to the mechanical properties this is quite obvious. The hardening curve of stress versus strain and the corresponding microscopic structures are the striking features of the mechanical properties of ductile materials. Despite the enormous effort that has been put into experimental and theoretical investigations, a completely satisfactory theory has not yet been attained. The principal reason for this is that a proper theoretical analysis of the highly nonlinear and nonequilibrium effects does not exist.

From a basic physical point of view a plastically deformed material is a highly nonlinear system that is driven far from thermodynamic equilibrium. Instability transitions and corresponding structural transitions are typical for these systems. To analyze plastic behavior, that is, nonlinear effects, two different approaches to microplasticity are currently under investigation: the theory of dislocations and continuum mechanics, respectively. The theory of dislocations has provided a thorough understanding of the behavior of individual defects and the interactions between them. However, it is doubtful whether this is a proper theory for describing a complex behavior of a dislocation structure under dynamic conditions, such as the formation of cells, veins, braids, or dipolar walls. The continuum approach, on the other hand, is complementary to the classical theory of dislocations and it is to some extent rather its counterpart. The advantage of a continuum mechanics approach is quite obvious: it may be very suitable for describing the global cooperative behavior of moving dislocations. This approach was initiatied by Holt [7], and extended by others [8, 9]. We will come to these points again in Section 1.5 of this Chapter.

In the following we take the classical theory of dislocations as a starting point since most TEM studies concentrate on differences in the hardness of various materials rather than explaining the complete hardening curve.

Theoretical descriptions of the strengthening of crystalline solids, as based on the classical theory of dislocations, involve the movement of dislocations against the resistance provided by various types of obstacles. The basic idea of strengthing is thus to introduce obstacles into the glide plane of dislocations. We may have one or more of the following obstacle types: forest dislocations, substitutional impurity atoms, and various kinds of precipitates. The strength of these obstacles depends on the exact nature and size of the obstacles. The resultant flow stress of the material can be calculated using the random obstacle theory for localized obstacles. In the case of coherent precipitates with a substantial misfit parameter, or when pressurized bubbles are considered, diffuse instead of localized forces may be used. Obviously, not only the nature and size of the obstacles but also their dispersion and concentration play important roles in such a calculation.

Undoubtedly, TEM has contributed a lot to the understanding of the physical nature and operative mechanisms of all these effects on the mechanical strength

of metals and alloys. However, the scientific questions that should be addressed by TEM remain open: in other words, does it really solve materials problems. Instead of giving a complete survey of the existing literature we intend to show some limited case studies within the realms and wealth of TEM in the field of metals and alloys. As stated before, when the scientific work is aimed at scrutinizing the structure–property relationship it would be of crucial interest to have suitable processing tools available to modify the structure as well, since structural features are also determined by the processing. It is well known that different heat treatments of metallic systems may produce various microstructures depending on the quenching rates applied. In our case a high-power carbon dioxide laser system turned out to be an excellent instrument for imposing heat treatments by which various structural features can be generated, depending on the interaction time with the specimen. Besides heat treatments, 'mechanical' impact by shot peening techniques and, on a more atomic level, by ion implantation are effective ways of altering the microstructural features. Both techniques have provided an impetus to the field of microscopy on metals and alloys.

Besides these more static investigations, that is, of the postmortem type, in situ deformations studies in TEM also made an impact to the understanding of the mechanical behavior. For that reason we include some observations on the interaction between dislocations and grain boundaries in ordered compounds by which we stress the materials investigations aspect of electron microscopy. The same point of view is taken regarding the structural aspects of boundaries investigated by means of high-resolution electron microscopy.

Finally, this Section is ended with an outlook and some conclusions.

1.2 Apparatus and Techniques for Imaging Metallic Systems

All electron microscopic observations presented in the following are carried out with either a JEOL 200CX microscope or a JEOL 4000EX/II ultra-high resolution microscope, unless otherwise stated. The general construction of a transmission electron microscope is discussed at length in part I of this series. An electrically heated tungsten or LaB_6 filament emits electrons, which are accelerated up to 200 keV. Focusing of the electrons is accomplished by using several electromagnetic lenses and electrical deflectors. For more information on the basic design of electron microscopes, reference should be made to Hirsch et al. [1]. The use of magnetic lenses makes it difficult to examine ferromagnetic materials, which interfere with the imaging magnetic fields. Generally the images obtained from such specimens are strongly astigmatic.

The most straightforward way to obtain information is to form an image of the transmitted beam. This is called bright-field imaging. Some of the electrons, however, are elastically scattered, leading to diffracted beams. In the normal bright-field mode diffracted beams are not allowed to pass through the objective aperture. The formation of an image of the sample is not the only way to obtain information. Another possibility is to use

the diffracted beams more explicitly. Some applications of electron diffraction are given by Hirsch et al. [1] and Thomas and Goringe [3]. This technique, that resembles other diffraction techniques, can provide information about the atomic structure of the sample. A diffraction pattern is obtained in the following way. For the directions for which the Bragg law, $n\lambda = 2d \sin\theta$, is satisfied, constructive interference of scattered electrons may lead to a diffracted beam. Here λ is the de Broglie wavelength, d_{hkl} the reflecting h, k, l plane distance, and θ is the half-angle between incident and scattered beam. The intensity of these beams is proportional to the geometrical structure factor [10]:

$$S_g = \sum f_j \exp(-2\pi i \boldsymbol{g} \cdot \boldsymbol{r}_j) \qquad (1)$$

where f_j is the atomic form factor and j runs over all the atoms of the basis, positioned at $\boldsymbol{r}_j$. Further, $\boldsymbol{g}_{hkl}$ denotes the reciprocal lattice vector, equal to $\boldsymbol{k}_0 - \boldsymbol{k}_i$, that is, the difference between the diffracted and the incident beams. The result of such a calculation is that for most crystals not all reflections are present. In an f.c.c. lattice, for instance, only those reflections for which h, k, l are all even or are all odd are present. In a b.c.c. lattice, only the reflections for which $h + k + l =$ even are present.

All beams for which the geometrical structure factor is not equal to zero (and the transmitted beam) are focused in the back focal plane of the objective lens. By suitable settings of the intermediate lenses, an image of this diffraction pattern can be obtained on the viewing screen or on a photographic plate. By placing an aperture in the image plane of the objective lens only that part of the specimen that is located inside of this aperture will contribute to the diffraction pattern. This technique, called selected area diffraction (SAD), is used to obtain certain information of a single grain or a large precipitate in a specimen, for example the crystal structure or the orientation. If a material is polycrystalline with a grain size much smaller than the selected area, many differently oriented grains contribute to the diffraction pattern. In this way a ring pattern, instead of a spot pattern, is obtained. It must be remembered that when correlating the SAD pattern to the image, all rotations caused by the magnetic lenses must be accounted for.

Making use of a diffracted beam instead of the transmitted beam is called dark-field imaging. Since the diffracted beams normally do not coincide with the optical axis of the microscope, the obtained image will not be of maximum quality, as far as spherical aberration is concerned. To overcome this problem the beam is tilted in such a way that the desired diffracted beam passes along the optical axis. Application of this technique leads to high-contrast images of various types of defects. Whenever an image of a lattice defect is formed by diffraction contrast, the specimen is tilted into a so-called two-beam condition. In this way only one set of diffracting planes contributes to the image formation, which makes interpretation of the micrograph more straightforward.

When dislocations are to be imaged, it is often better to tilt the crystal slightly further by such an amount that the exact Bragg condition is only fulfilled within a small region near the dislocation. The deviation from the exact Bragg condition for a perfect crystal is then given by $\boldsymbol{s} = \boldsymbol{k}_0 - (\boldsymbol{k}_i + \boldsymbol{g})$. In this

way a high-resolution image is obtained in which the dislocation shows up as a white line. This is called weak-beam imaging [11].

The amount of extra tilting can be determined accurately by the relative position of the so-called Kikuchi pattern with respect to the diffraction pattern. The origin of this Kikuchi pattern lies in the elastic rescattering of inelastic scattered electrons. It is almost independent of the incident beam direction but seems to be connected to the lattice. The position of the diffraction spots, however, is determined by the incident beam direction. As a consequence, the exact deviation from the Bragg condition is specified, and a numerical value of s can be obtained.

The magnitude of s is related quantitatively to the resolution of a dislocation image in the following way. The width of a dislocation image is approximately $0.3\xi_g$. Here ξ_g is the extinction distance for a certain reflection, defined for $s = 0$ (non-relativistic):

$$\xi_g = \frac{\pi V \cos\theta}{\lambda s_g} \tag{2}$$

and ξ_g is of the order of 15–200 nm for most metals. V is the volume of the unit cell and θ is the scattering angle for the specific reflection. The width of a dislocation imaged in bright field or dark field (i.e., $s_g = 0$) can be detrimental to observations of, for example, the separation between two dislocations that are very closely spaced. In the case of a tilt away from the Bragg condition, the periodicity of the oscillation is calculated from two-beam dynamical theory, in which it is assumed that the incident wave amplitude ϕ_0 and the diffracted amplitudes ϕ_g both vary with depth, z:

$$\xi_g^{\text{eff}} = \frac{\xi}{\sqrt{1 + w_g^2}} \tag{3}$$

where $w_g = s_g \xi_g$. From this equation it can be concluded that the effective extinction distance decreases for increasing deviation away from the Bragg condition. If a large deviation from the Bragg position ($s_g > 0.2\,\text{nm}^{-1}$) is established, the effective extinction distance decreases according to Eq. (3), and a much sharper dislocation image is obtained. This technique is called weak-beam diffraction. The width of a dislocation image can be reduced to values of the order of 1–5 nm.

Planar faults, such as stacking faults, will have a displacement $\boldsymbol{R} = 0$ above the fault and $\boldsymbol{R} = \boldsymbol{R}_f$, the fault vector, below the fault. If the fault is inclined to the foil surface, an oscillation in the intensity of the fault image is produced, unless $\alpha = 2\pi \boldsymbol{g} \cdot \boldsymbol{R}_f$ equals $n2\pi$. In this case no phase change is introduced, and the fault will be invisible. Application of the kinematical theory for electron diffraction, which assumes that the crystal is so thin or s so large that the intensity of the diffracted beam is small compared to the intensity of the transmitted beam, results in the following expression for the amplitude of the diffracted beam:

$$\phi_g = \frac{\pi \mathrm{i}}{\xi_g} \phi_0 \int_0^t \exp(-2\pi \mathrm{i} s_g z)\, \mathrm{d}z \tag{4}$$

where t is the thickness of the foil. The kinematical theory is qualitatively quite successful but is not quantitatively reliable since it assumes that ϕ_0 is constant with depth z. It becomes clear that sometimes unphysical results may be obtained, such as for small s and large $t|\phi_g(t)|^2$ exceeds

$|\phi_0|^2$. This can be concluded from the evaluation of the integral, which yields the following expression for the diffracted intensity as a function of foil thickness (assuming $\phi_0 = 1$):

$$I_g = \frac{\pi^2 \sin^2(\pi t s)}{\xi_g^2 (\pi s)^2} \tag{5}$$

provided that s is large, that is, the kinematical theory is applicable. It is easy to see that the diffracted intensity thus oscillates as a function of depth, giving rise to the so called thickness fringes. These fringes, occurring with a periodicity ξ_g^{eff}, can be used to estimate the foil thickness. Such an estimate must be done carefully, because a small deviation from the Bragg condition already results in a significant effect on the extinction distance. Therefore, the value obtained must be regarded as an upper limit rather than an exact value.

Contrast of lattice imperfections is described by a modified version of Eq. (4) [12]:

$$\phi_g = \frac{\pi i}{\xi_g} \phi_0 \int_0^t \exp[-2\pi i(s_g z + \boldsymbol{g} \cdot \boldsymbol{R})] \, dz \tag{6}$$

where $\boldsymbol{R}$ describes the displacement of a particular atom from the perfect lattice, in the vicinity of an imperfection, and can be calculated using (an-)isotropic linear elasticity theory [13, 14]. Although a precise calculation of the image contrast of, for instance, a dislocation is quite complicated, this equation allows one to predict the visibility of a dislocation. For the most simple case, the pure screw dislocation, we know that $\boldsymbol{R}$ has the same direction as the dislocation Burgers vector, $\boldsymbol{b}$. Therefore, all contrast of a screw dislocation vanishes when $\boldsymbol{g} \cdot \boldsymbol{b} = 0$, and the dislocation is invisible. General dislocations usually have a dislocation displacement field with more components, and now the conditions for invisibility become $\boldsymbol{g} \cdot \boldsymbol{b} = 0$ and $\frac{1}{8} \boldsymbol{g} \cdot \boldsymbol{b} \times \boldsymbol{u} \leqslant 0.08$, where $\boldsymbol{u}$ is the dislocation line vector.

Equation (6) can also be used to calculate the contrast of coherent precipitates, since the lattice is often distorted by such a particle. Semi- and incoherent precipitates, showing little or no strain contrast, can best be imaged by making a dark field image using a precipitate reflection. This leads to high-contrast images of bright precipitates in a dark matrix. Apart from these two sources of contrast, precipitates may also become visible through absorption differences or by the presence of Moiré and thickness fringes [1].

In the case of voids and bubbles, image contrast is formed in several ways [15]. Absorption contrast will dominate when the bubble diameter forms a significant fraction of the foil thickness. In this case a bubble will appear as a bright object. When diffraction contrast plays a role, a bubble can also appear as a dark object, strongly depending on s [16]. Highly pressurized bubbles deform the host lattice, and the strain field can be imaged. Bubbles also give rise to phase contrast, which can be used to enhance the contrast. This is carried out by defocusing the image [17]. In an under-focus condition the bubble appears as a bright area with a dark periphery, whereas in an over-focus condition the contrast is reversed. The diameter of the bright area in an under-focus condition is generally regarded as the real diameter.

Sometimes image and diffraction information is not sufficient to identify the nature of a certain precipitate. This may

be the case since lattice parameter measurements using TEM are not very accurate, and additional information can be obtained by energy dispersive spectrometry (EDS) [18]. In this technique, the electron beam is focused onto the precipitate to be analyzed. Energy, absorbed by sample atoms, is partly released by emitting X rays. From the X-ray spectrum the chemical composition can be derived.

1.3 Metallic Specimen Preparation for Transmission Electron Microscopy

Because of the high electron absorption of the samples, the maximum sample thickness should be of the order of a few hundred nanometers. Thinner samples, however, result in an improved image quality. To obtain such a thin foil, which is also sufficiently rigid to be handled, requires the following procedure. First a thin sheet (100–200 μm thickness) of the material to be examined is cut. When a laser-treated material is to be examined, this usually means that the sheet is cut plan parallel to the treated surface. This is done by using a low-speed diamond saw with low mechanical distortion. Then from this sheet, disks (3 mm diameter) are produced using a spark cutter or a mechanical puncher. Subsequently, these samples are thinned from one or both sides in such a way that in the center of the disk the removal rate is slightly higher than away from the center. As soon as an optical system detects a hole in the center of the disk, the process is stopped. In this way, a wedge-shaped electron transparent region comes into being at the edges of the hole. The actual thinning can be accomplished in two ways, as described below.

1.3.1 Electromechanical Thinning

With this technique, the specimen is thinned by the removal of material in an electrochemical process. This technique should be preferred over other techniques because mechanical and thermal distortion are almost negligible. The polishing process is performed while keeping the sample under a positive electrical voltage with respect to a polishing solution. An extensive list of electropolishing solutions has been published by Kelly et al. [19]. For most iron alloys a solution of 10% perchloric acid, 85% acetic acid, and 5% water suffices. For aluminum alloys other specific solutions are necessary. The electrical current is adjusted in such a way that a viscous layer is obtained at the sample surface. Only then polishing instead of etching takes place.

The formation of a viscous layer is also favored at low temperatures. Development of gas bubbles at the surface should be avoided since this leads to irregular surfaces. These can be avoided by lowering the temperature or the electrical voltage. Apart from the sample and the electrolyte, a light source and a sensor are used. When the sensor detects light, the sample is sufficiently thinned and the polishing process is stopped; the sample is then removed and rinsed in alcohol and methanol.

This process works best in the case of a bulk sample consisting of a homogeneous single-phase material. In multiphase

materials there is nearly always the problem of preferred polishing. Preferred polishing is also often observed just near the edges of laser tracks, and is caused by structural and compositional changes at these points.

When an ion-implanted material is to be examined, thinning must be done from one side only, the bulk side. The implanted surface is covered with wax so that only the bulk side is in contact with the polishing solution. Often, when the sample is perforated, the polishing solution creeps between the wax and the edge of the perforated foil, thereby etching away the implanted layer. To overcome this and the aforementioned problem with the laser-treated materials, another thinning technique is also applied, namely ion milling.

1.3.2 Ion Milling

An ion milling machine for TEM specimen preparation was first constructed by Castaing [20]. Nowadays, milling machines, such as the Gatan Dual Ion Mill Model 600 used in our research, are constructed. The sample is mounted on a rotating sample holder in a vacuum chamber. Two ion guns are pointed at the sample center under a low incidence angle (5–15°), from both the top and the bottom. With this technique, thinning takes place by sputtering caused by the incident focused ion beam. As a rule, argon ions are used, because of their inertness and their low solubility in metals. Additionally, they have a good damage to sputtering ratio. The sputter rate can be further optimized by working in a good vacuum, in order to prevent continuous formation of surface layers such as oxides and hydrocarbons. The milling process is stopped when a laser optical system detects a hole in the specimen.

Prior to ion milling the samples are treated with a mechanical dimpler (Gatan Model 656 dimple grinder in our research), in order to reduce the milling time or to select the depth of the electron transparent region. This is accomplished by making shallow dimples on one or both sides of the specimen. The sample is mounted on a rotating table, while a rotating grinding wheel covered with a diamond slurry removes small amounts of material. This process is carried out under a small load and with a low speed until a thickness of approximately 40 μm is reached.

Implantation damage is the main disadvantage of the milling technique, and can become visible as small argon bubbles or as clustered point defects. Furthermore, thermal damage can be a problem because the heat conduction in thin foils is evidently not rapid. This complication plays a role especially in the case of low melting point metals such as aluminum. To circumvent this problem, the sample holder can be cooled down to liquid nitrogen temperature. An optimum must be found between the sputter rate and specimen damage. In practice, this means that the gun voltage and current are set to 3–4 kV and 1 mA, respectively, while the angle of incidence is 10–15°.

The main advantage of this specimen preparation technique is that almost every material can be prepared. Only metal–ceramic composites may give rise to problems because of a large difference in sputtering yield. The most important advantage is that the electron transparent region can be positioned quite accurately.

This is particularly important for laser-treated samples, where the microstructure may change when going from the center to the edge of the track.

1.4. Typical Examples

1.4.1 Observations of Static Dislocations by Transmission Electron Microscopy

As said before, the leading theme of this Section is to show how electron microscopy may be used to relate structural information to the physical properties of metallic materials. Let us concentrate first on the microstructural analyses of metallic systems heat treated by a less conventional technique such as laser processing, and see how transmission electron microscopy reveals the relevant physical mechanisms in enhanced mechanical properties. Generally, a fine solidification structure and an enhanced solid solubility form the basic concepts behind improvements by laser melting. Further, freezing-in of metastable phases, creation of metallic glasses, or even formation of quasicrystalline phases are also feasible.

Apart from these advantages, there also exist disadvantages: during cooling, the metal surface will shrink locally, usually leading to substantial stresses. Sometimes these stresses will cause severe cracking. In other cases the tensile stress will only lead to a decreased behavior of wear or fatigue. To overcome the stress-related problems, the combination with ion implantation has been explored [21, 22]. In addition, another novel approach was invented: laser melting combined with a shot peening treatment afterwards [23, 24].

While it is in use it is often the surface of a workpiece that is subjected to wear and corrosion. Recent research has been devoted to the modification of only the surface of a relatively cheap material, instead of making the entire workpiece from a higher grade of metal. In view of the practical implications, the development of wear-resistant surfaces has been taken as a starting point in this work.

As a general rule, wear is determined by the interplay of two opposing properties: ductility and hardness. Wear can be reduced by modifying the surface layer in such a way that it acquires higher ductility, so that greater plastic deformation can occur without particles breaking off. Soft surface layers can be very effective in reducing wear due to delamination. Resistance to wear by abrasion, on the other hand, is then low.

However, wear can also be reduced by making the surface layer harder. Then again, increasing hardness also means an increase in the elasticity limit and a reduction in ductility, leading to a lowering of fatigue resistance and hence to brittle fracture. The characteristics of the system (i.e., whether the wear is caused by delamination or abrasion) determine which of these two methods should be chosen.

Over the past few decades, ion implantation has been promoted for the modification of surface layers of materials. Machines for ion implantation have so far been developed primarily for the semiconductor industry. Although the improvement of the mechanical properties of ion-implanted surfaces is also known to be considerable, it should not be assumed

that the technique has been fully perfected for tribological applications. Commonly used ions in the field of ion implantation metallurgy are those of nitrogen, carbon, boron, titanium, molybdenum, and chromium, where the first species are mainly used to form second phase particles and the last ones form a surface alloy. The mechanisms leading to improved surface properties are: initiation of a more favorable wear mechanism, increasing the hardness through the formation of second-phase particles or through solid solution hardening, and surface alloying that may improve the corrosion resistance.

As the penetration depth is inversely related to the ion mass, at constant accelerating voltage, light elements are favorable. In this context one should also mention the recent development to increase the implantation energy, up to the megaelectronvolt regime [25], and the mechanism of so-called tribo-enhanced diffusion [26]. The latter mechanism applies to nitrogen implantations, where it has been reported that the layer with increased wear resistance sometimes extends up to several micrometers, which is an order of magnitude deeper than the original implantation depth. Analysis showed that nitrogen was still present in the near-surface layer after a long period of surface layer removal by wear.

The basic idea behind our approach of applying ion implantation to laser-treated metals is to build up a compressive stress that might be able to reduce or even reverse some unfavorable properties such as tensile stresses and retained austenite of laser-treated materials. For a better understanding, these two topics will be investigated from the TEM viewpoint to see whether this basic idea is right or wrong.

1.4.1.1 Atomic Peening

RCC steel (2.05 wt.% carbon, 11.05 wt.% chromium, 0.62 wt.% tungsten, and the balance iron) is very suitable for laser treatment when a constant hardness profile is required [27]. Because a martensitic transformation is absent, the tensile stresses are remarkably high after laser treatment [28]. The dislocation density is also high, but not extremely so, 10^{12}–$10^{14}\,m^{-2}$. After laser melting, a dendritical solidification structure develops, consisting of austenitic cells surrounded by M_3C carbides. Before implantation the samples were smoothened by grinding with SiC paper followed by polishing with a 3 μm diamond paste.

Implantations were carried out with doses of 3×10^{16}, 1×10^{17}, and $3 \times 10^{17}\,Ne^{+}\,cm^{-2}$ at 50 keV. Figure 1 shows a TEM image of an implanted foil, exhibiting the depth distribution of the bubbles throughout an austenitic cell. A distribution of the bubble radii is shown in Fig. 2. The mean bubble radius is 20 nm and the mean volume is $4.5 \times 10^{4}\,nm^{3}$.

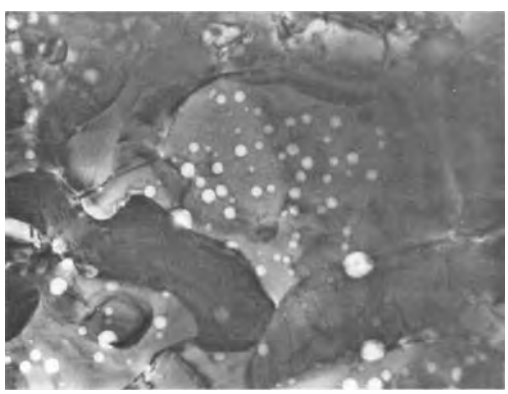


Figure 1. TEM image of neon bubbles in laser-treated RCC steel (stereo angle 20°, dose $3 \times 10^{17}\,Ne\,cm^{-2}$).

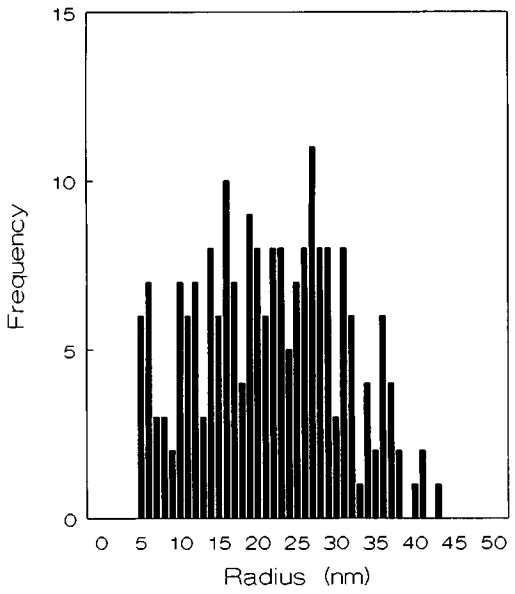


Figure 2. Bubble size distribution after implantation of 3×10^{17} Ne cm^{-2} into laser-melted RCC steel.

The bubble density amounts to 5.7×10^{20} bubbles m^{-3} [21].

One of the remarkable findings by TEM is that the stereo measurement of the depth distribution reveals that the mean implanted range is approximately twice the distance expected. This might be explained by channeling of neon during the first stage of implantation. The mobility of neon, which gives rise to the observed bubble formation, may also contribute to an extension of the implantation profile. Furthermore, it is known that large stresses, which are present in these samples, may enhance the diffusion [29, 30]. Finally, a high dislocation density may also promote diffusion by a pipe diffusion mechanism.

The mechanical properties of laser-melted RCC steel are significantly improved by neon implantation [22, 31], and the question is whether TEM can shed some light on the relevant mechanism. It turned out that (Fig. 3), in contrast to the original idea of pressure-induced hardening, this improvement is likely to be due to the dislocation–bubble interaction. Dislocations may bypass the bubbles either by pure glide (Orowan process) and cross slip, or by shearing the bubbles. The stress field around a bubble determines whether a dislocation will reach the bubble and penetrate or will bypass by bowing around the bubble. Usually, in aged-hardened alloys the contribution of modulus hardening to the shearing force, required to cut a particle by glide dislocations, is small. Since the elastic energy of a dislocation is proportional to the shear modulus, a change of energy and hence a force will be associated with a dislocation interacting with a particle whose shear modulus differs from that of the matrix. The extreme case is the interaction with a void or a bubble, in which the elastic energy is reduced to zero and the system can be described by image forces. So, the modulus hardening in the case of neon bubbles may be a crucial contribution to the hardening, since in the neighborhood of the bubbles the dislocations always experience an attractive force component towards the bubbles, which is confirmed by TEM observations.

In Fig. 3a the critical shear stress [32] is depicted as a function of bubble size, given a volume fraction of 0.025. In this graph, the Orowan stress is calculated using the traditional Orowan expression, rewritten for spherical obstacles of radius R and a fixed volume fraction [33]. Pressure hardening is calculated, combining an expression for shearable objects [34], and an expression for the average force between an edge dislocation and a

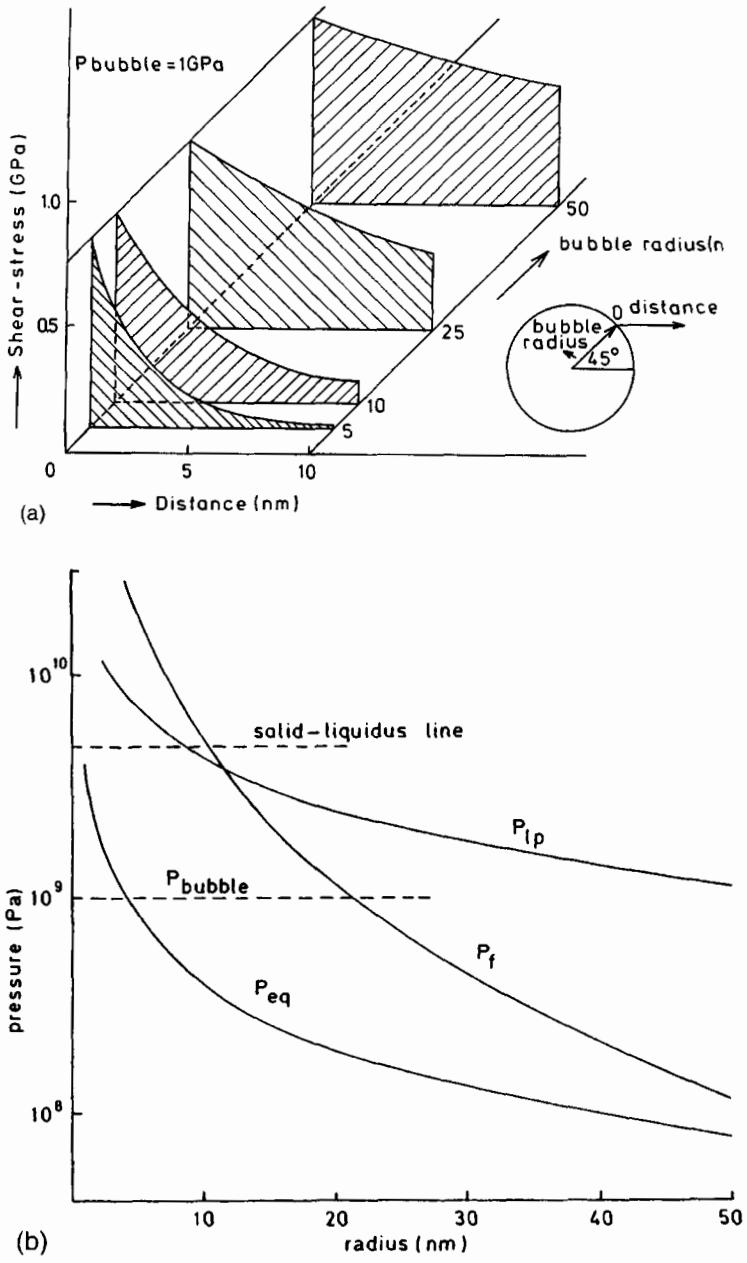


Figure 3. (a) Shear field of a bubble as a function of distance and size of the bubble. (b) Bubble pressure versus bubble radius. P_{eq} represents the equilibrium pressure, P_f is the pressure for interbubble fracture, and P_{lp} represents the minimum pressure for loop punching.

pressurized bubble [35, 36]. Finally, modulus hardening is calculated using an expression quoted by Hanson and Morris [37]. It is easy to see that modulus hardening contributes considerably compared to Orowan and pressure hardening. Apart from the hardening caused by the interaction of neon bubbles and dislocations, the reduction of the tensile stresses will also contribute to the

improved wear behavior. Still another possible cause of the improved wear behavior may be formed, namely by the local stress field around the individual bubbles, which might influence the work hardening rate.

1.4.1.2 Stress Fields in the Implanted Layer

During implantation, Frenkel pairs are produced; about 450/neon atom at 50 keV. Most of them annihilate, but neon will be trapped by the remaining vacancies. As a consequence the interstitials, forming loops or trapped at dislocations, will contribute to a compressive stress. This stress field can be of the order of 1 GPa [38, 39]. When a neon–vacancy complex absorbs an interstitial neon atom or coagulates with another bubble, the pressure in the new bubble may exceed the equilibrium pressure ($P_{eq} = 2\Gamma/r$, where Γ is the surface free energy; $\Gamma = 2.5\,\mathrm{J\,m^{-2}}$ is often used for f.c.c. iron [39]). This excess pressure can be reduced by emitting interstitials or dislocation loops, thereby contributing to a compressive stress. The bubbles have a stress field consisting of hydrostatic and shear components [40]:

$$\sigma_H = \frac{\bar{p} f_v}{1 - f_v} + \frac{(p - \bar{p})\alpha f_v}{1 - \alpha f_v}$$

$$\tilde{\sigma}_{ij} = \frac{R^3}{2r^3}\left(\delta_{ij} - 3\frac{x_i x_j}{r^2}\right) \times \left(\frac{\bar{p}}{1 - f_v} + \frac{p - \bar{p}}{1 - \alpha f_v}\right) \tag{7}$$

where f_v is the volume fraction of the bubbles. r and x_i represent the length and components of the radius vector with the origin in the center of the bubble of radius R. δ_{ij} is the unit tensor and α is defined by

$$\alpha = \frac{1 - \bar{G}/G}{1 + 4\bar{G}/3K} \tag{8}$$

where $\bar{p}$ and $\bar{G}$ represent the average pressure and average shear modulus, respectively. The idea is that around each bubble a cell is constructed and that this cell is embedded in an effective medium with elastic constants $\bar{G}$ and $\bar{K}$, containing all other bubbles with an average pressure of $\bar{p}$. However, in general the individual pressure p of a bubble will differ from this average pressure. It is easy to see that σ_H plays a minor role because it scales with f_v. Now the question is whether for the TEM data we can derive an estimate of the pressure in these bubbles, P_{eq}. We know already the equilibrium pressure as a function of the radius. Suppose the thermal vacancy concentration is rather low, then the bubbles may grow by punching out dislocations. For large dislocation loops of radius R_l, the minimum pressure for loop punching is simply

$$p_{lp} = \frac{2\Gamma}{r} + \frac{Gb}{2\pi(1 - \nu)R_l}\ln\frac{R_l}{r_{eh}} \tag{9}$$

where r_{eh} is the equivalent hole radius (including core energy effects). With increasing bubble concentration, inter-bubble fracture may take place. For a cubic array of bubbles, the pressure needed for fracture is [41]

$$p_f = \frac{2\Gamma}{r} + \sigma_f\left(\frac{1}{\pi r^2 C_b^{2/3}} - 1\right) \tag{10}$$

where σ_f is the fracture stress and C_b is the bubble concentration. An estimate of the fracture stress might be the theoretical fracture stress of cleavage fracture. The

interbubble distance is considered to be large [42]. A more appropriate value is the strength needed to overcome the lattice resistance to dislocation motion, which is approximately 0.001 G for f.c.c. metals. In the present case, p_f might be somewhat larger because of the rather flat distribution of bubble radius, and small as well as large bubbles coexist. In Fig. 3b the pressure is presented according to these predictions. To a first approximation the bubble pressure is at least equal to the equilibrium pressure for small bubbles. Therefore, TEM predicts for the bubble pressure 1 GPa (dashed line in Fig. 3b).

The shear stress can be calculated as a function of bubble size at various distances from the edge of the bubble. This is illustrated in Fig. 3a, assuming a realistic bubble pressure of 1 GPa. When there are two bubbles close to each other the fields will accumulate and may lead to interbubble fracture when the interbubble distance becomes small. When on the other hand the interbubble distance becomes too large, there is no repulsive force on emitted dislocation loops, and the pressure can more easily be lowered to equilibrium pressure.

Figure 4 shows TEM micrographs of an SS304 sample implanted with $3 \times 10^{17}\,Ne^{+}\,cm^{-2}$, and prepared following the bulkside-thinning method. Here, bubble radii up to 40 nm are observed within a broad distribution. From the diffraction patterns and the dark field micrographs it became evident that the dark areas represent a b.c.c. phase. In contrast to other reports in the literature [43, 44] the b.c.c. phase has hardly any orientation relationship with the matrix. It is also noteworthy that the size of the particles is smaller than that observed previously [43]: 50 nm instead of 120 nm.

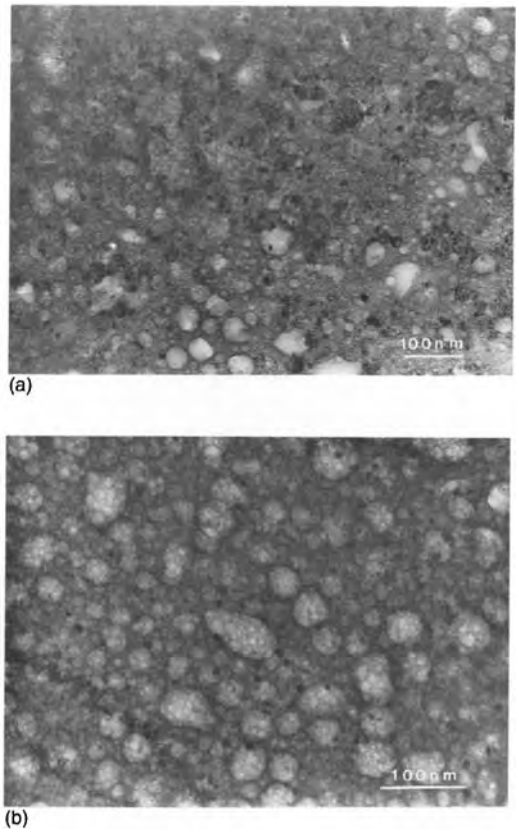


Figure 4. (a) Bright field transmission electron micrographs of an implanted bulkside thinned sample (dose $3 \times 10^{17}\,Ne^{+}\,cm^{-2}$) and (b) an implanted thin foil, showing bubbles and b.c.c. particles.

In a thin foil, implanted with $1 \times 10^{17}\,Ne^{+}\,cm^{-2}$, the bubble radii vary from 1 to 5 nm, and no b.c.c. phase can be detected. When a thin foil is implanted with $3 \times 10^{17}\,Ne^{+}\,cm^{-2}$, the bubble size distribution is the same as in the bulkside-thinned sample, but the b.c.c. particles look more spherical and are even smaller: about 8 nm in diameter (Fig. 4b). In this case there is no orientation relationship at all. The critical dose

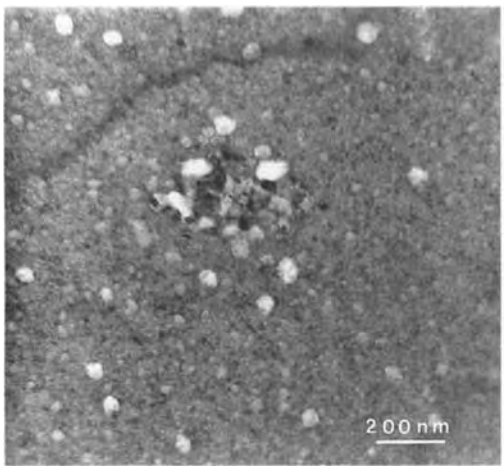


Figure 5. Bright field transmission electron micrograph showing b.c.c. particles in the vicinity of bubbles, grown after an anneal (dose 1×10^{17} Ne^+ cm^{-2}).

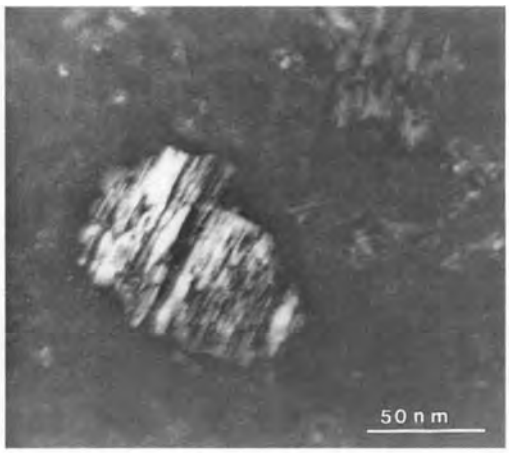


Figure 6. Dark field micrograph of a large twinned b.c.c. particle.

for transformation in thin foil samples amounts to about 1.5×10^{17} Ne^+ cm^{-2}.

To investigate the effect of the bubble size, an implanted thin foil, dose 1×10^{17} Ne^+ cm^{-2}, has been annealed at 400°C for 144 h. After this anneal the maximum bubble size has increased from 5 to 25 nm. Figure 5 shows that in the vicinity of a cluster of bubbles, a few b.c.c. particles have grown.

In the annealed thin foils, twinning is frequently observed, e.g. Fig. 6 shows a b.c.c. particle which is larger than average and internally twinned in a sample annealed for 67 h. The twins are 1–4 nm in size and are observed in the bulkside thinned samples as well.

Analysis of the experimental results suggests that a shear stress produced at the neon bubble is a plausible nucleation site for the martensitic transformation. Although the magnitude of the stress field is large, its range strongly depends on the bubble size. The critical dose of about 2.3×10^{17} Ne^+ cm^{-2} (in the bulk samples), is exactly in the range where there is a strong increase in bubble size.

Beside the stress fields of the bubbles there is also a radiation damage-induced stress field. Assuming that a biaxial compressive stress of 1 GPa exists in the top layer, this means that apart from a hydrostatic component of 670 MPa, a shear stress of 500 MPa acts on two planes. It is unlikely that this stress is the main reason for the transformation, because especially in the thin foil experiments an efficient phase transformation was observed, whereas it seems impossible that such large stresses can exist in a thin foil. On the other hand one must realize that in a thin foil a martensitic transformation easily occurs because of the large free surface. By inspecting stereomicrographs it was not possible to obtain a three-dimensional impression of the b.c.c. particles because of the random orientation and thus random Bragg conditions. Therefore, it is not possible to conclude whether the phase transformation occurs preferentially at the bubble surface or in between two bubbles due to the

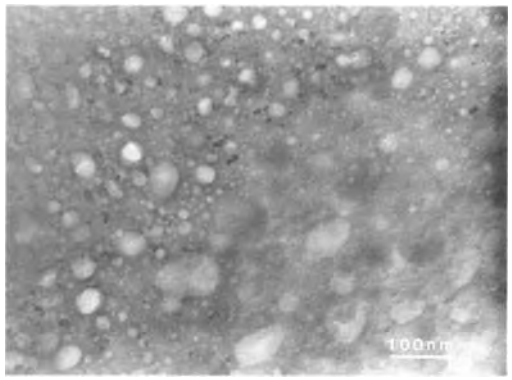


Figure 7. Bright field micrograph of a bulkside thinned sample showing no b.c.c. phase at the edge.

accumulation of the stress fields. The empty region of the edge of the bulk-side thinned samples, which represents the top of the implanted layer, supports the conclusion that the transformation does not start at the surface, but deeper down where there is a higher bubble density (Fig. 7). Preferred nucleation was observed several times at grain boundaries in thin foil samples implanted below the crucial dose (Fig. 8). Observation in an off-Bragg condition reveals a coarsening of the bubbles along the grain boundary. It is likely that this coarsening is caused by an enhanced diffusion process along the boundary. The large bubbles at the boundary may cause the transformation, but it is also likely that the nucleation energy at the grain boundary is lower anyway [45]. It cannot be concluded which factor is predominant.

From the annealed specimens we prove that the bubble size is a crucial factor in the martensite nucleation. It is important that the bubbles in the annealing experiments maintain their pressure. This means that it is not allowed to raise the temperature to a value where thermal vacancies become available, because in that case the bubbles might grow by absorption. According to Evans [46], one should keep the temperature below $0.45T_m$, which is satisifed in our experiment. So it is reasonable to assume that the large bubbles in the annealed samples still are at least pressurized to the equilibrium pressure of the small bubbles from which they are grown. If coarsening takes place, then there are two possible explanations. First, the gas in the bubbles may go back into solution, re-emerging preferentially into larger bubbles where the pressure might be lower. Secondly, the small bubbles themselves may migrate as entities, collisions between bubbles may follow, leading to coalescence and hence coarsening. There is no evidence

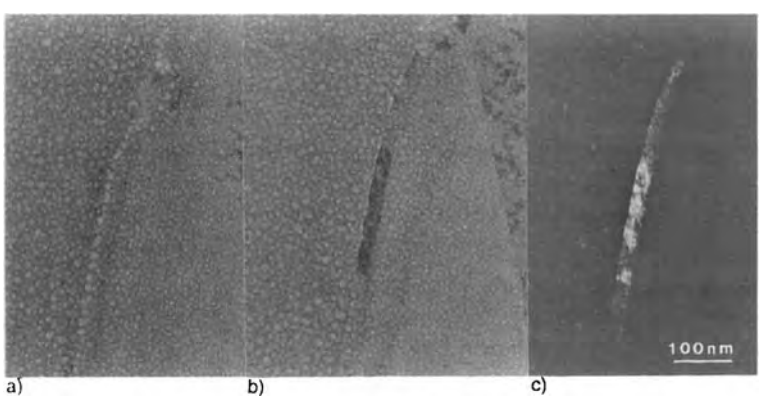


Figure 8. (a) Bright off-Bragg condition, (b) Bright in-Bragg condition, and (c) dark field micrographs of α-phase at a grain boundary in a thin foil sample (dose 1×10^{17} Ne^+ cm^{-2}).

in favor of the first alternative. The difficulty of it lies in the high energy of solution (about 5 eV for neon in iron) rather than the migration energy. Taking the second alternative, neon bubble migration can be expected in a temperature gradient, and due to interaction with dislocations during annealing. Since bubble migration requires the transfer of solid material from one side of the bubble to the other, it seems likely that surface diffusion around the bubble periphery occurs upon annealing at 400°C. Further, when dislocations move by climb during annealing one may expect them to sweep up the bubbles in their path, and hence cause a coarsening. In addition, Goodhew [47] pointed out that around the annealing temperature of $0.45T_m$, surface diffusion is the predominant mechanism of coarsening. The growth rate appears to increase with increasing bubble concentration. In the present study the bubble concentratton is very high, that is, 10^{24} bubbles m^{-3}.

The main difference between our result and the ones obtained previously [43, 44] is the lack of a distinct orientation relation. However, it has to be emphasized that Johnson et al. [43] refer to antimony implantations, implying that the stress fields are not localized around bubbles, and to krypton as an implant [44]. Since neon possesses a much higher compressibility compared to krypton, it is quite obvious that the critical dose of neon implantations should be higher. Consequentially, the nucleation behavior is altered in such a way that the volume fraction and bubble sizes are much larger in our case. This, together with the high pressure in the bubbles, causes a deformation around the bubbles, and a reduction of the maximum size of the nuclei. The small size together with the deformation causes the random orientation. This is in agreement with the increase in the orientation relationship when we compare the unannealed thin foil samples (3×10^{17} Ne^+ cm^{-2}), that is, substantially deformed with the bulk-side thinned (3×10^{17} Ne^+ cm^{-2}) or annealed thin foil samples (1×10^{17} Ne^+ cm^{-2}), as can be concluded from Fig. 9. In the case of the bulk-side thinned samples there is a restriction to the deformation because there is only one free surface. In the case of the annealed samples the volume fraction of the bubbles is much lower, causing less deformation. The variation in sizes of the b.c.c. particles supports this idea.

The interesting outcome from the TEM observations is that, primarily, noble gas bubbles have been shown to be detrimental for wear resistance, unless they play an active role in, for example, pinning of dislocations or favoring an f.c.c. to b.c.c. transformation, as observed in implanted 304 steel. The austenite stability is not low enough in the case of laser-treated RCC steel, since no transformation has been observed in the implanted samples. Nevertheless, one expects an increased work hardening rate in such a layer. Concerning the f.c.c. to b.c.c. transformation, we can state that the shear field of an inert gas bubble may induce a local martensitic transformation, provided the bubble size and pressure are large enough. Further, the austenite stability must be low.

1.4.1.3 Micro-Peening

Shot peening is a process of cold working applied to the surface of metals by the

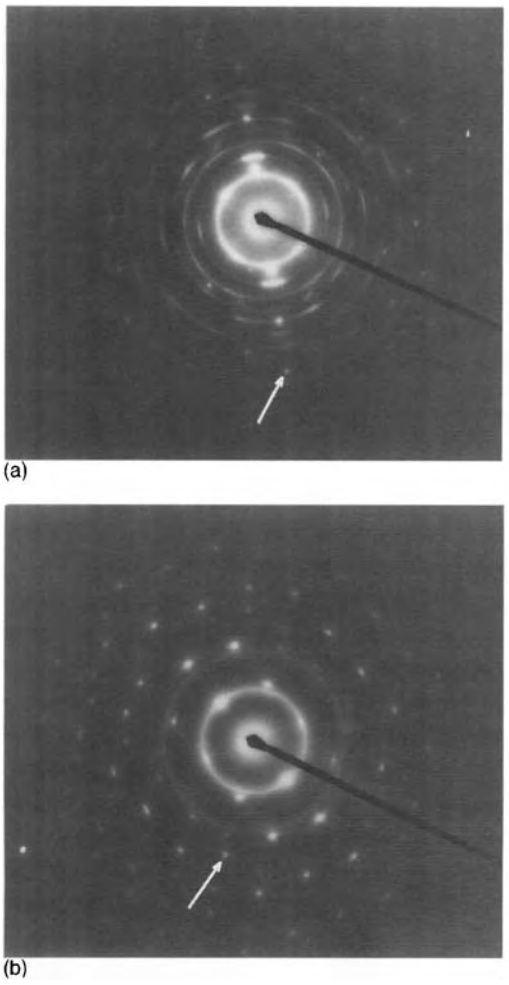


Figure 9. Diffraction patterns of implanted samples, showing (a) angular intensity dependence of the b.c.c. rings in the case of the bulkside thinned sample and (b) a constant intensity in the case of the thin foil sample. In both patterns a ⟨400⟩-type f.c.c. reflection is indicated that can be used for indexing the b.c.c. rings.

impingement of a stream of particles under controlled conditions. Although this is a recent definition, similar processes such as hammering have long been used [48]. Modern techniques, for example robotics, make a controlled and reproducible process feasible. It is even more important that modern analyses of peened surfaces provide a feedback control, so that the process parameters can be optimized. In the past decades this has provided a new impetus to shot peening.

The principal aim of shot peening is to improve the fatigue lifetime, stress corrosion resistance, and other tensile stress-related metal failures. These improvements are achieved mainly by the compressive surface stress state that arises by this treatment. In such a layer it is more difficult for existing cracks to propagate and to nucleate. This compressive stress also acts as an off-set when a workpiece is loaded in such a way that the resultant surface stress state would be a tensile one. Because shot peening is capable of inverting a tensile stress state to a compressive one it is an appropriate treatment of a surface after grinding, machining or any other treatment that caused the surface to be in a tensile stress state. It must be noted that, while a workpiece is in service, stress relaxation will take place. Therefore, repeening may be useful to restore resistance to fatigue failure [49].

Apart from changing the stress state, shot peening may also be applied to clean, form and strengthen metal surfaces [50]. The reduction of wear rate after shot peening of steel is reported by Jiawen [51] and Ritchie [52]. Strengthening of a metal surface will also have an effect on the fatigue behavior. Usually the higher dislocation density, as introduced by peening, has no beneficial effects. In softer alloys the improvement of the fatigue behavior is mainly due to strain hardening [53].

It has already been stressed by other researchers that not only the peening conditions but also the microstructural state of the material is crucial with respect

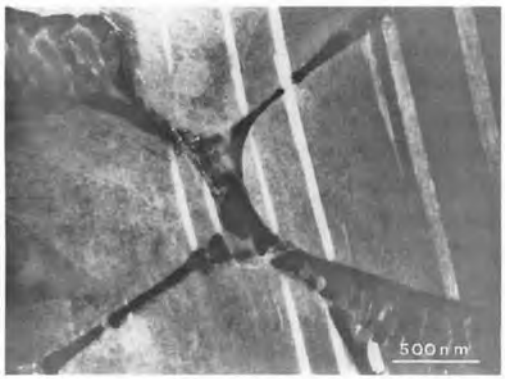


Figure 10. Dark field transmission electron micrograph showing strain-induced martensite in austenitic cells.

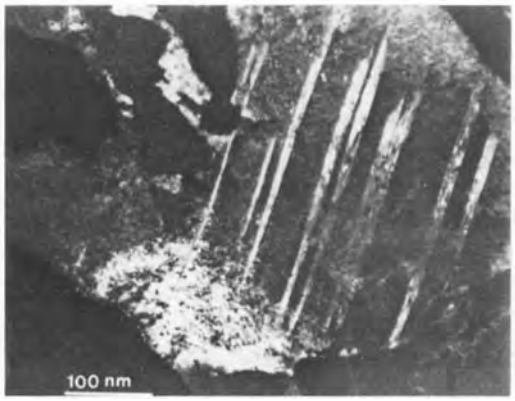


Figure 11. Wear induced α-martensite near the surface.

to the final results [54]. They studied the mechanical properties of two aluminum alloys, peened at different stages of the heat-treating process. Their results showed strikingly different resultant properties with changing tempering conditions. It also became clear that no general conclusion can be drawn, since the optimal treatment depends on the composition of the alloy as well as on the requested specific property. An interesting TEM observation is displayed in Fig. 10, showing an RCC sample after shot peening. TEM specimens are prepared using a dimpler followed by ion milling in such a way that various depths below the surface can be inspected (Fig. 10). At depths of at least 15 μm, stress-induced martensite is found with the same morphology as observed in a worn laser-treated RCC sample (Fig. 11). At a depth of 30 μm only an increase in dislocation density was found: $10^{13}\,m^{-2}$ before and more than $10^{14}\,m^{-2}$ after peening.

Theoretical descriptions of the shot peening process are not readily available. The stress fields that arise from the multiple impact are quite complicated, which makes application of work hardening theories complex. Therefore, research in the field of shot peening is still characterized by its empirical nature. It is likely that this situation will not change very quickly unless more TEM research is carried out. So far, because of the analogy with Hertzian stress analysis, the best results are obtained in predicting residual stress profiles.

Next, two different examples are presented of aluminum alloys, shot peened and investigated by TEM afterwards. In the past decades many aluminum alloys have been designed to serve various purposes. Binary Al–Si alloys, for instance, show superior cast properties, but have a low strength. Alloys containing copper, magnesium and small fractions of other alloying elements, on the other hand, possess poor castability, but develop high strength after appropriate heat treatments. Improvement of these alloys by conventional processes has its limits, and therefore alternative techniques such as splat cooling and laser melting are increasingly being explored. The surfaces are given a shot-peening treatment to improve the

residual stress state, and the resultant microstructure is investigated with TEM. Here we will concentrate on a eutectic Al–12 wt.% Si (Al–12 Si) alloy representing the class of cast alloys, whereas Al 2024 was chosen to represent the commercial Al–Cu–Mg class of alloys.

One of the reasons for choosing an Al–Si alloy is its simple phase diagram, that is, no intermetallic compounds. Further, eutectic Al–Si alloys are commonly used for automotive parts such as cylinder heads and pistons. These types of alloys combine, apart from good castability, low density, good corrosion resistance and low thermal expansion. The mechanical properties can be further improved by small additions of sodium or strontium [55]. These impurities cause a change in the morphology of the silicon particles, and a decreased spacing between them. The amount of modifier, however, is quite critical since over-modification results in reduced mechanical properties. Furthermore, losses from the melt and the formation of undesired intermetallic compounds cause additional problems.

Rapid solidification by means of laser surface melting offers another way of achieving a fine distribution of the silicon phase [56]. The cellular solidification results in cell sizes of the order of 1 μm, depending on the scan velocity. The depth of the laser tracks is 50–500 μm, which is enough to allow further machining, and depends on the process parameters. In addition, laser surface treatment is a process that offers greater precision and less distortion than most other surface techniques.

Comparison of the optical micrograph of Al–Si in the as-cast condition (Fig. 12) with the transmission electron micrograph

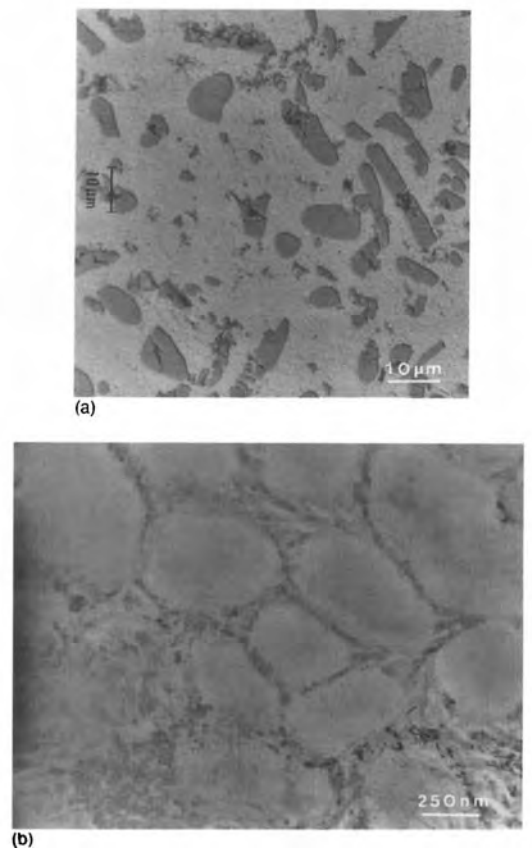


Figure 12. (a) Optical micrograph of aluminum–silicon in the as-cast condition and (b) transmission electron micrograph of the cell structure after laser melting at 8 cm s^{-1}.

of the cell structure after laser melting (Fig. 13) clearly reveals the refined structure caused by the rapid solidification after laser treatment. Transmission electron micrographs are used to estimate the dislocation density of the samples before and after shot peening. Because the observed densities are without exception very high [56] (see Fig. 13), only rough estimates can be given here: 10^{13} m^{-2} for a low laser scan velocity without shot peening and more than 10^{14} m^{-2} for a high laser scan velocity after 360 s of shot peening. Dislocation cell

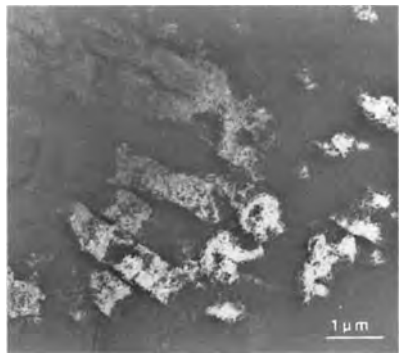


Figure 13. Dark field transmission electron micrograph showing the dislocation structure in peened Al–12 Si (laser scan velocity 8 cm s^{-1}).

structures are not observed to be formed during the shot peening. Apparently, the sizes of the cells to start with are too small to allow further division. Minimum energy configurations by the formation of a checkerboard-like structure of the original cells are sometimes observed (Fig. 13).

An important TEM observation is the difference in mobility of the dislocations in the low and the high laser scan velocity samples. In the $\frac{1}{8}$ cm s^{-1} samples the heat input from the electron beam was sufficient to cause thermal activated dislocation movement. In the 8 cm s^{-1} sample this was not observed.

The low laser scan velocity samples reveal a breaking up of the solidification structure after shot peening. This change of structure is not accompanied by an increased dislocation density, which implies dynamic recovery during deformation. The result is a fine granular structure with a typical grain size of 0.5–1 µm (Fig. 14). Silicon particles originally forming the eutectic cell walls are then more randomly distributed.

Another observed phenomenon in high laser scan velocity samples after shot peening is the formation of silicon precipitates.

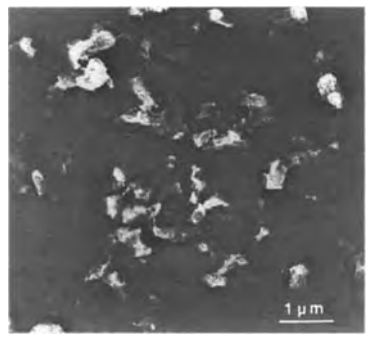


Figure 14. Dark field transmission electron micrograph of the fine granular structure in a sample with a laser scan velocity of $\frac{1}{8}$ cm s^{-1}, after shot peening.

The sizes of these precipitates vary from approximately 5 to 50 nm (Fig. 15).

Most important is the breaking of the original silicon particles in the peened bulk samples (Fig. 16). This indicates that this alloy in the as-cast condition is not suitable for a shot-peening treatment.

In this section an estimate is presented of the contribution to the flow stress originating from the obstacles present in our samples. The following relevant obstacles are present: substitutional silicon atoms, forest dislocations and silicon precipitates. The strength of these various types of obstacles can be calculated,

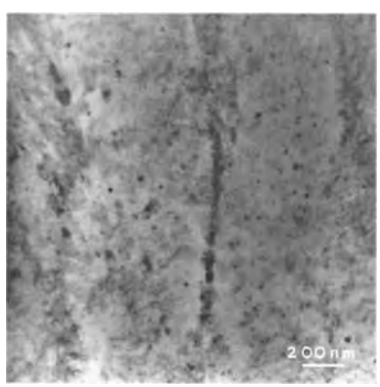


Figure 15. Precipitates in shot-peened Al–12 Si (laser scan velocity 8 cm s^{-1}).

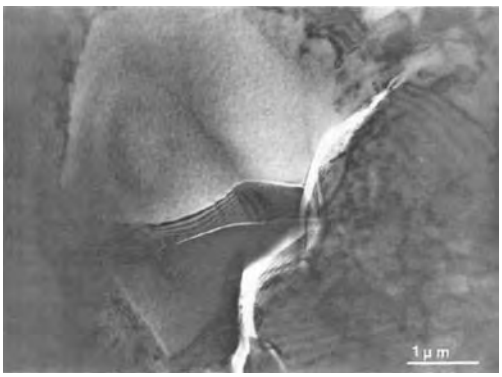


Figure 16. Shot peening induced breaking of primary silicon particles in the as-cast Al–12 Si alloy.

approximating the flow stress by

$$\tau = \tau_{\text{subst}} + \tau_{\text{disl}} + \tau_{\text{prec}} \tag{11}$$

where the first two contributions can be described using Friedel's relation [23]:

$$\tau_{\text{subst,disl}} = \frac{Gb}{L} \cos^{3/2}\left(\frac{\phi_c}{2}\right) \tag{12}$$

where G, b, L, and ϕ_c are the shear modulus, magnitude of the Burgers vector, obstacle spacing and critical break away angle, respectively. In the material the relatively high amount (up to 2%) of silicon in solid solution is likely to play an important role. The maximum solubility of silicon in aluminum is only 1.5% at the eutectic temperature, and it decreases to 0.05% at 300°C [57]. This low solubility is caused by the large misfit between aluminum and silicon, while silicon still occupies substitutional sites. This large misfit (18%), however, gives the silicon atoms a large interaction strength, and thus an already large contribution to the flow stress at relatively low concentrations. To calculate the numerical value of this contribution an estimate of the interaction strength ξ is made by comparing ξ with misfits in other systems [58]. ξ depends on the critical break away angle being $\xi = \cos(\phi_c/2) = 0.04$. This value, together with L rewritten as $b/\sqrt{c}$, results in a contribution to the flow stress of $10^{-3}G$, at a solid solution concentration of 1.5%. This value of the flow stress has to be considered as an upper limit since a justification of Friedel's relation can be given only for not too weak obstacles and for a low concentration of impurities, that is, when the mean distance between obstacles is larger than the minimum curvature radius of the dislocation at the obstacle. In fact the critical condition of Friedel's approximation [59], $cW^2/\xi b^2 < 1$, indicates that the present situation is a borderline case (W represents the width of the dislocation). A different approach as presented in Mott–Nabarro's model [60], taking into account the internal stress around each solute [61, 62], predicts a contribution of the flow stress of $4 \times 10^{-4}G$.

Regarding the misfit between aluminum and silicon, one should note that silicon is a nonmetal and that the tabulated atomic radius might change significantly when embedded in a metal. The rate of lattice contraction, when it is added as a solute, indicates that the misfit is smaller than that based on the atomic radii.

In addition, Fleischer [62] pointed out that the modulus mismatch is just as important as size mismatch when considering solid solution hardening. To take modulus hardening into account one should have a knowledge of the relative change of the shear modulus G^{-1} $(\delta G/\delta c)$ as solute is added. Since silicon is virtually insoluble in aluminum this change is very difficult to measure, and to our knowledge no values have been reported in the

literature. Nevertheless, the modulus mismatch will have its effect on the interaction strength. Therefore, we have made an approximation of this mismatch by applying the following relations [63]:

$$\frac{(G_{Al}+G_l)f_v}{G_{Al}+G_l+(G_{Si}-G_{Al})(1-f_v)} \leqslant \frac{G-G_{Al}}{G_{Si}-G_{Al}} \leqslant \frac{(G_{Al}+G_g)f_v}{G_{Al}+G_g+(G_{Si}-G_{Al})(1-f_v)} \tag{13}$$

where, if $(G_{Si}-G_{Al})(K_{Si}-K_{Al}) \geqslant 0$,

$$G_l = \frac{3}{2}\left(\frac{1}{G_{Al}}+\frac{10}{9K_{Al}+8G_{Al}}\right)^{-1}$$

$$G_g = \frac{3}{2}\left(\frac{1}{G_{Si}}+\frac{10}{9K_{Si}+8G_{Si}}\right)^{-1}$$

In these equations $G_{Al,Si}$ and $K_{Al,Si}$ represent the shear and bulk moduli of the aluminum and silicon phases, respectively, and f_v denotes the volume fraction of silicon. It must be emphasized that this method neglects all electronic effects, and should only be used as a first-order approximation.

The resulting value of the relative modulus change and the value of the change in the lattice parameter can be combined using the method developed by Fleischer [62] in order to obtain a numerical value for the interaction strength. This method results in $\xi = 0.042$, which is very similar to the estimate based upon the nonembedded atomic radius.

The second contribution to the flow stress, arising from the dislocation–dislocation interaction, also needs to be considered. The numerical value of the dislocation density can be obtained by TEM. The estimates based on the transmission electron micrographs are of the order of $10^{13}\,m^{-2}$ after laser melting, and are raised by one or two orders of magnitude after subsequent shot peening. To calculate the numerical value of the contribution to the flow stress, Friedel's equation is rewritten as

$$\tau_{disl} = \alpha G b\sqrt{\rho} \tag{14}$$

where α is $\frac{1}{3}$, and ρ represents the dislocation density. This equation is valid if internal stresses are absent. In heavily deformed materials the internal stress takes up about 80% of the applied stress [64]. Therefore, the effective contribution to the flow stress amounts one-fifth of the value calculated with the equation above, resulting in a value of τ_{disl} of $2 \times 10^{-4}G$ for a dislocation density of $10^{14}\,m^{-2}$.

During shot peening, the measured concentration of silicon in solid solution significantly decreases, and silicon precipitates appear. This low-temperature precipitation must then be assisted by pipe diffusion. Because the capture radius of the dislocations is undoubtedly small compared to the initial dislocation spacing, a lot of deformation is needed. This is confirmed by TEM observations. The presence of these silicon precipitates form the third contribution to the flow stress. Since these precipitates are incoherent they will not be sheared. Then the critical shear stress is given by the Orowan stress [65]:

$$\tau = \frac{0.84Gb}{2\pi(1-\nu)^{1/2}}\frac{\ln(2R_s/r_0)}{(L-2R_s)} \tag{15}$$

where R_s represents the particle radius and r_0 is the dislocation core radius.

To calculate the contribution of the precipitates to the flow stress, their average

spacing is needed. This is a difficult task because the small precipitates in particular are easily overlooked, due to the random orientation of the precipitates and the heavily deformed condition of the material. Therefore, only an estimate of 0.1 μm is used here. This value results in an Orowan stress of $4 \times 10^{-3}G$.

Comparison of the calculated contributions to the flow stress indicates, based on TEM analyses, that the high hardness values after laser melting and shot peening are caused mainly by precipitation hardening and, to a lesser extent, by solid solution hardening. The dislocation–dislocation interaction seems not to be the predominant factor either.

The considerations mentioned above do not include any possible changes in cross-slip behavior. We have not yet discussed the large difference in dislocation density in samples treated at low and high laser scan velocities, as observed by TEM. Examination of the experimental results indicates that a changed cross-slip behavior might be responsible for this phenomenon.

An increase in hardness can be achieved if one could eliminate or postpone the onset of cross-slip (stage III). From that point on, the work-hardening rate is decreasing with increasing strain. In our samples we see that after an initial period of approximately linear hardening, the work-hardening rate indeed decreases, but the peening time at which the saturation starts increases with increasing laser velocity. This indicates that the onset of cross-slip might be postponed by an increased laser velocity. This observation is in line with the observed higher dislocation density in samples treated at high laser scan velocity. In the present case the cross-slip behavior can be affected in various ways.

First, the introduction of impurity atoms affects the stacking fault energy and therefore the stress necessary to constrict the partial dislocations to cross-slip. In the case of aluminum the stacking fault energy is of the order of $200\,\mathrm{mJ\,m^{-2}}$. This high value results, applying linear elasticity, in an equilibrium separation less than one Burgers vector for a screw dislocation. Since in our samples the maximum amount of silicon impurities is 2%, it is not very likely that the stacking fault energy is lowered by such an amount that the cross-slip behavior changes significantly.

A second mechanism by which, in our case, the cross-slip behavior can be altered has to do with the prerequisite of the cross-slip process to occur, namely that the dislocation is in the pure screw orientation over some distance. When impurities with sufficient stress fields are present, the dislocations will show deviations from straightness, and the activation energy for cross-slip will apparently be raised.

A study of the coarse slip bands after deformation may reveal whether this second mechanism plays a dominant role or not. Small additions of copper or magnesium in aluminum, elements that cause a significant misfit, have been shown to increase the stress for the onset of cross-slip [66] whereas addition of silver has no such effect. Since there is almost no size difference between aluminum and silver, this is in accord with the second mechanism described above. As already mentioned, there is a large misfit in the case of silicon in aluminum, so that, in our case, it is likely that cross-slip is hindered by this mechanism (see Fig. 17).

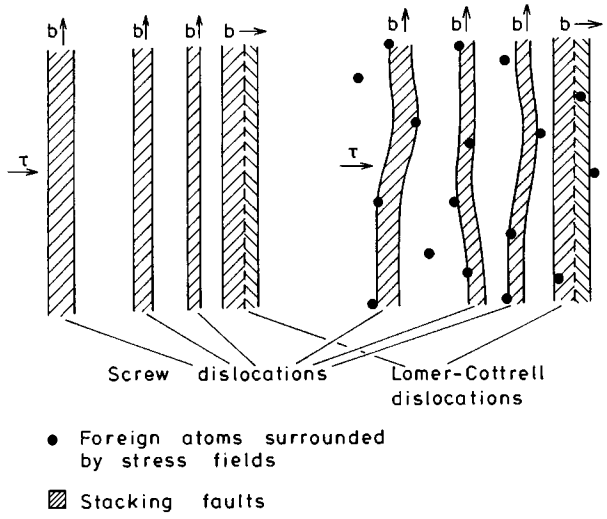


Figure 17. The inhibition of cross-slip by foreign atoms interacting with dislocations.

It has been found that laser treatment of an as-cast eutectic Al–Si alloy leads to a very fine dispersion of the silicon phase in an aluminum matrix. Further, it turned out that this laser treatment is able to amplify considerably the effectiveness of a subsequent shot-peening treatment. In particular, the maximum attainable hardness and compressive stress increase upon increasing the quench rate, that is, upon increasing the laser scan velocity. From TEM observations the creation of shot-peening induced silicon precipitates forms the most striking result of this study. It is likely that this precipitation is caused by the capturing of silicon atoms by moving dislocations. The resultant distribution of small silicon precipitates make up the predominant contribution to the flow stress.

Another example concentrates on aluminum alloys containing copper, magnesium and small weight percentages of other alloying elements which develop high strength after appropriate heat treatments. Conventional hardening treatments of these aluminum-base alloys usually exploit the beneficial effects of a high concentration of precipitates. In binary Al–Cu alloys that form the basis for all alloys in the 2XXX series, usually containing 4 wt.% of copper, the precipitation of intermetallic compounds is known to proceed through a series of different precipitates. Normally, ageing starts with fully coherent Guinier–Preston (GP) zones, which can initially be regarded as copper disks, one atomic layer in thickness. As the GP zones grow, they lose coherency in the direction normal to the disk. This gives rise to GP-II zones or θ'' precipitates. These platelets, approximately 0.5 nm thick and 10 nm across, lie parallel to the {100} planes in the aluminum matrix. On raising the temperature these platelets transform into other semi-coherent platelets, designated as θ'. These precipitates have the same orientation relationship with the matrix, but are an order of magnitude larger than the θ'' platelets. The series terminates with the incoherent equilibrium phase θ. The elimination of the internal stresses associated with the semicoherent

precipitates and the accompanying coarsening are responsible for the softening observed when the precipitation sequence reaches this stage.

If the alloy also contains magnesium, the Al–Cu precipitates are partly or completely replaced by Al–Cu–Mg precipitates. These precipitates are also platelets, and are called S′ and S in the literature. The composition, crystal structure and orientation relationship of these precipitates depend on the copper/magnesium ratio. With magnesium in excess of 4 wt.%, freezing of an Al–4.4 wt.% Cu alloy may follow the Al–S branch leading to the so-called T phase (approximately $CuMgAl_6$). If no copper is present, Mg_2Al_3 will be formed.

Here, laser hardening of an Al 2024 alloy was carried out to investigate the possible hardening mechanisms in laser-treated Al–Cu–Mg alloys. Normally, hardening by laser treatment is achieved primarily by the small grain sizes that are formed as a result of the rapid solidification process. The flow stress varies as $d^{-\alpha}$ where α depends on the cell size and cell wall type. Since the cell size is inversely related to the solidification rate, generally an increasing hardness is observed with increasing laser scan velocity. The nucleation and growth of precipitates that would contribute predominantly to the hardness of laser-melted materials is usually suppressed by the high cooling rates involved, typically of the order of 10^4–10^5 K s^{-1}. In an Fe–W–C tool steel, however, the presence of precipitates in a surface that was 'autotempered' by applying overlapping laser tracks was reported [27]. Nevertheless, in overlapping laser tracks, precipitation processes are not likely to occur.

In aluminum–copper alloys, however, the precipitation may start with (semi-) coherent platelets which significantly lowers the nucleation energy. Therefore, in such alloys precipitation may be a feasible process. The scan velocity was varied over two orders of magnitude to obtain various stages in the precipitation sequence. Subsequent hardness measurements were carried out, before and after the shot-peening treatment, to study the work-hardening behavior of the laser-treated material. In addition, attention was paid to the microstructure and the residual stress state in the samples.

Samples of commercial Al 2024-T3 (4.4% copper, 1.5% magnesium, 0.6% manganese and the balance aluminum; solution treated, cold rolled and naturally aged) were given the same laser treatment as the Al–Si samples. For this material,

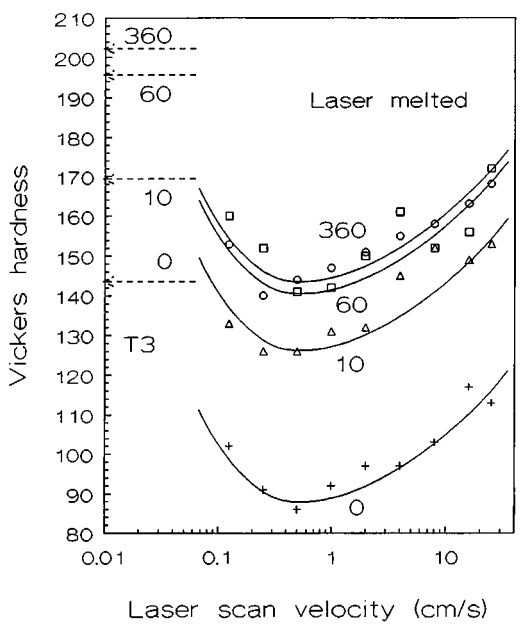
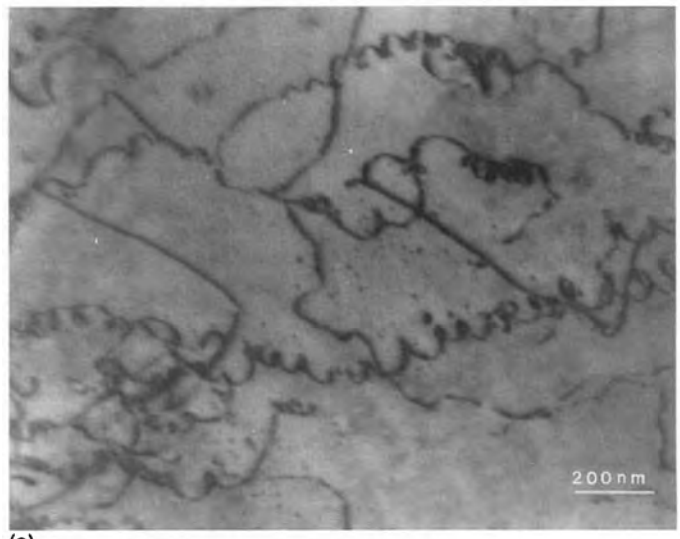
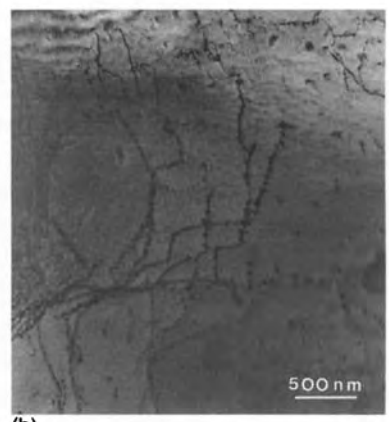


Figure 18. Hardness profile as a function of laser scan velocity and shot-peening time (seconds).



(a)



(b)

Figure 19. Helical dislocations in a sample treated at a laser scan velocity of $\frac{1}{8}$ cm s^{-1}: (a) grown by vacancy absorption and (b) by interaction of a screw dislocation with prismatic vacancy loops. Schematic representations of the operating mechanisms are depicted in (c) and (d), respectively.

however, there are no homogenization problems. Therefore, the scan velocity range was extended up to 25 cm s^{-1} (single pass). It should be emphasized that of all the elements in Al 2024, magnesium has by far the lowest boiling point (1390 K), and it is therefore to be expected that a substantial fraction of magnesium will actually evaporate during the laser treatment. Some extra samples of pure aluminum, Al–2.32 wt.% Cu and Al–4.5 wt.% Cu, were also given a laser treatment. Subsequent shot peening was carried out only for the Al 2024 samples. TEM investigations are used to explain the remarkable behavior of the mechanical properties of this alloy as displayed in Fig. 18.

TEM specimens taken from Al 2024 samples and exposed to different laser scan velocities, before shot peening, show marked differences in microstructural features. The samples treated at lower scan velocities, $\frac{1}{8}$ to $\frac{1}{2}$ cm s^{-1}, are mainly precipitate-free, although in some areas

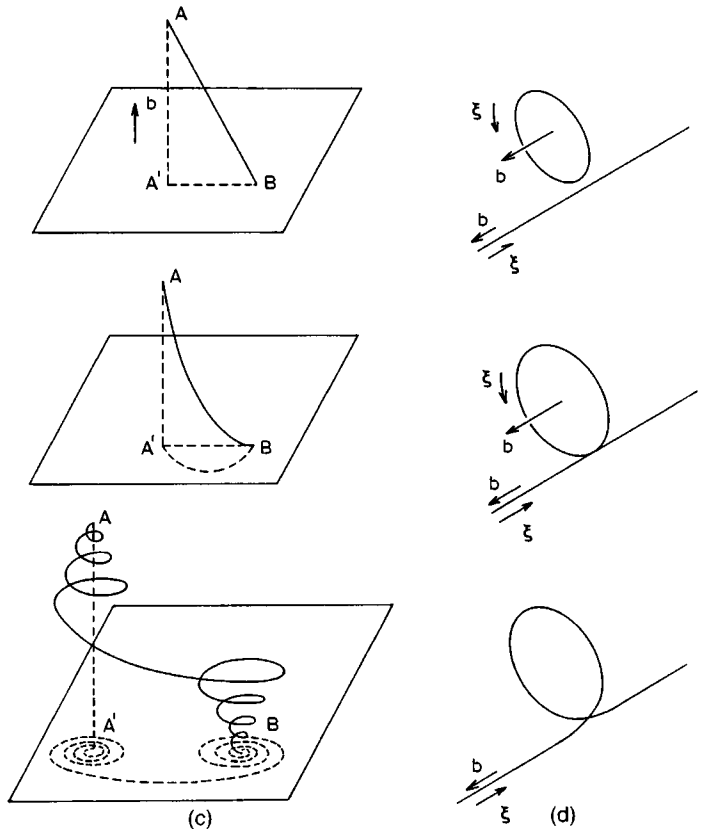


Figure 19. *Continued*

coarser precipitates could be observed. However, it is uncertain whether these are partly dissolved cell boundaries or real precipitates.

The most striking feature is the formation of helical dislocations (Fig. 19a and b), as observed in all samples treated at low laser scan velocity. Most of these dislocations exhibit a uniform equilibrium shape with respect to pitch and radius. At the lowest scan velocity, however, non-uniform shapes are occasionally observed (Fig. 20).

In samples treated at laser scan velocities of between 1 and 4 $cm\,s^{-1}$ the precipitation of the tetragonal platelets, with an average size of 150 nm, is observed, as can be seen in a bright field micrograph (Fig. 21). Since these platelets are oriented parallel to the {100} faces of the aluminum matrix, we may conclude that these precipitates are not the so-called S′ precipitates (Al_2CuMg). The S′ precipitates grow locally on {120} planes, giving {110}, {100} or {130} as an overall growth plane [67, 68]. In addition to the platelets a more globular precipitate is also observed in these samples, mainly nucleating on the edges of the θ' plates and on dislocations. This could be the θ phase as well as the S or S′ phase. Since in alloys with a copper/magnesium ratio of less than 8/1, Al_2CuMg precipitates can start to play a role [69], the existence of the S and S′

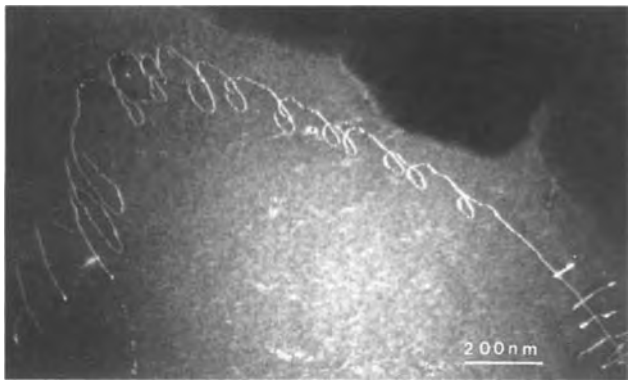


Figure 20. Dark field transmission electron micrograph of a nonequilibrium helical dislocation.

phases cannot be excluded. Electron diffraction patterns have not yet revealed the nature of these precipitates.

Samples treated at laser scan velocities of 8–25 cm s^{-1} show only a cellular structure, in which it is seen that the cell boundaries act as obstacles for dislocation motion. GP zones, which could precede the formation of the θ' plates, are not observed.

After shot peening, in all samples a higher dislocation density was observed by TEM. Apart from the increased dislocation density, however, precipitates could now be detected in samples that did not develop observable precipitates before shot peening. A typical example of this precipitation process induced by shot peening in a sample treated at a low laser scan velocity of $\frac{1}{8}$ cm s^{-1} is shown in Fig. 22.

The concentration of precipitates seems also to be higher than before peening in samples treated in the scan velocity range of 1–4 cm s^{-1}. This might indicate an increased solid solution in these samples, but the smaller average size, 100 nm, will also result in a higher concentration. Again, most of these precipitates are θ' platelets, aligned with the cubic axes. A second type of precipitate, smaller in size but also of the shape of a platelet, is identified as the S′ phase, initially growing on [120] planes. These precipitates have a

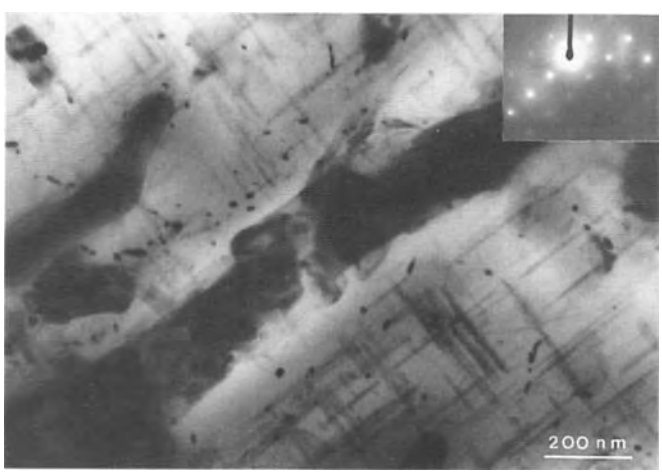


Figure 21. Bright field transmission electron micrograph showing θ' platelets and some incoherent θ or S precipitates, in a sample treated at 2 cm s^{-1}.

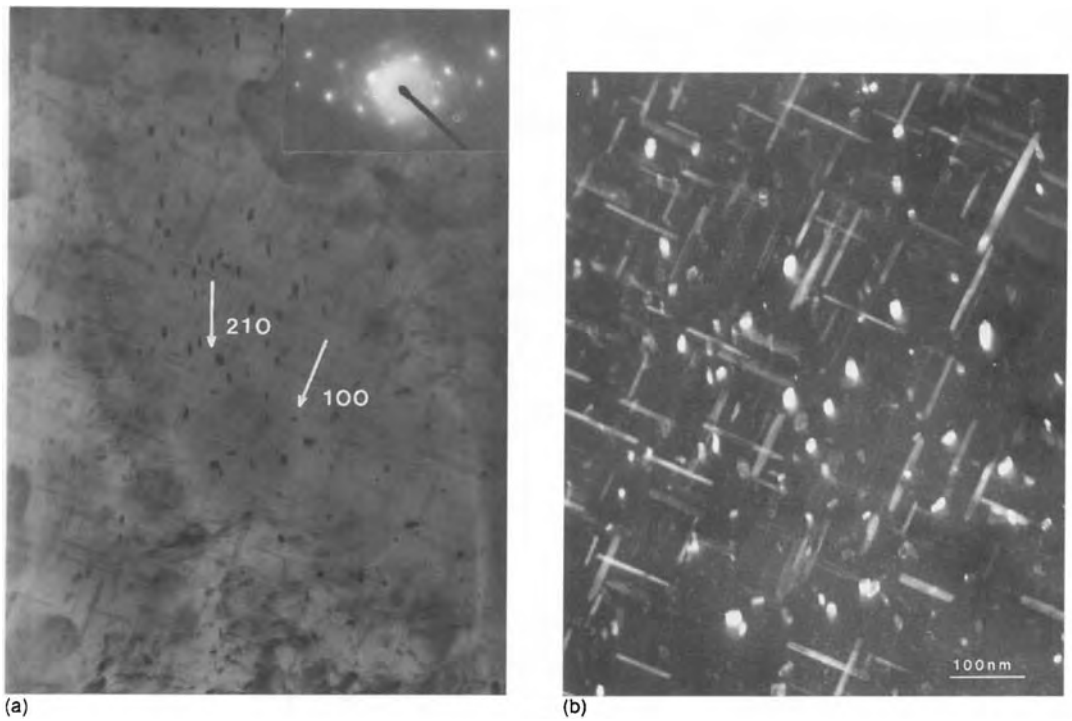


Figure 22. (a) Bright and (b) dark field transmission electron micrographs of shot-peening induced θ' and S' precipitates in a sample treated at a laser scan velocity of $\frac{1}{8}$ cm s^{-1}, peened during 360 s. The crystallographic planes on which the precipitates are observed to grow are indexed.

much smaller width-to-thickness ratio and will therefore contribute less to an increase of the flow stress than the θ' plates (at the same volume fraction). The very broad size distribution, down to only 2 nm, makes this contribution more uncertain.

To discuss the TEM results and to see how they can contribute to the explanation of the hardness profile, the laser scan velocity range is divided into two regimes: the low scan velocity regime ($\frac{1}{8}$ to $\frac{1}{2}$ cm s^{-1}) and the high scan velocity regime ($\frac{1}{2}$ to 25 cm s^{-1}).

Low scan velocity regime

Since the cooling rates are lower in the low laser scan velocity samples, compared to 2 cm s^{-1} samples, in which precipitation of the θ' plates is observed, one would expect to see an increasing nucleation and growth of precipitates in these samples. In contrast, below 1 cm s^{-1} no significant precipitation is observed. The explanation for this observation lies in the decreasing vacancy concentration with decreasing laser scan velocity. Apparently at the lower laser scan velocity the vacancies are annealed out before the temperature reaches values that would favor the nucleation and growth of observable precipitates, a.o. by absorption at dislocations causing the formation of helical dislocations. The rise in hardness values in the low-velocity regime cannot be caused by precipitation hardening

but must be due to two other contributions.

First, the dissolution of the cell walls and subsequent homogenization will play an important role. Since there exists only a small temperature interval during which homogenization takes place, as can be determined from [70], the cooling rate must be low enough to allow sufficient time for diffusion. The copper (and magnesium) brought into solid solution will contribute to an enhanced hardness, but for reasons described above, copper will not form any precipitate. The dissolution may continue as long as not all cell walls have been fully dissolved, provided that the cooling rate is low enough. Obviously this situation has not been reached in practice, and it explains why the hardness is still increasing with decreasing laser scan velocity.

In the additional experiments performed on the alloy with 2.32 wt.% of copper, it was observed that the cell walls were completely dissolved, only near the edges of the low scan velocity ($\frac{1}{8}$ cm s^{-1}) laser tracks. This observation is illustrated in Fig. 23. From this experiment we conclude that to a first approximation only half of the copper is in solid solution in Al 2024 treated with a scan velocity of $\frac{1}{8}$ cm s^{-1}. A calculation of the diffusion distance in a $\frac{1}{8}$ cm s^{-1} track, during cooling in the interval between 660 and 450°C, applying the analytical model for the temperature profile [71], yields a distance of approximately 0.1 μm. A calculation for magnesium yields a similar value. These values, which should only be regarded as being a first approximation, are too small for significant homogenization. Nevertheless, it shows that homogenization may play a role only at the lowest scan velocities.

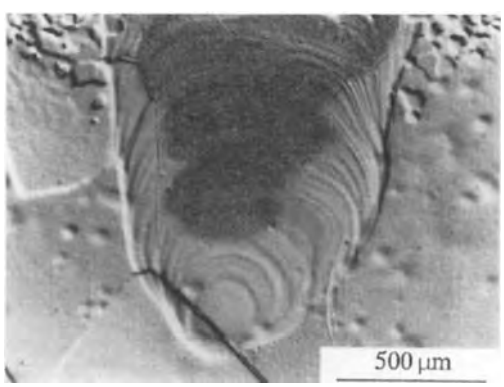


Figure 23. Dissolved cell walls in a $\frac{1}{8}$ cm s^{-1} track on an Al–2.32 wt.% Cu sample.

To quantify the contribution of solid solution to the hardness the same methods as in the previous section dealing with Al–Si are applied, taking size as well as modulus mismatch into account. This calculation yields for copper and magnesium the same critical break away angle, for which $\cos(\theta_c/2) = 0.036$. In the case of copper this high value is mainly due to modulus mismatch, whereas in the case of magnesium, size mismatch contributes most. The corresponding contribution to the flow stress amounts to 19 MPa at 2.32 wt.%.

Besides homogenization, a second contribution to the increased hardness in the low scan velocity regime is related to the formation of helical dislocations. Since the glide of a helical dislocation is restricted to its cylindrical surface on which it is wound, these dislocations are essentially immobile. The formation of helical dislocations is observed in all samples treated at low laser scan velocity.

From the geometry of these dislocations it is concluded that they arise from both climb of the part of a mixed

dislocation of edge character (see Fig. 19a and c) and from the interaction with prismatic vacancy loops (see Fig. 19b and d). Both mechanisms require a supersaturation of vacancies [13, 72]. A large quantity of vacancies is present because of the substantial cooling rates. The reason that the formation is possible only in samples treated at the low laser scan velocity is that only here is the cooling rate at intermediate temperatures low enough to allow for adequate vacancy diffusion. Once the temperature is too low for significant vacancy diffusion, the dislocations become immobile over the distance that has been affected by climb or by the interaction with the vacancy loop. The equilibrium shape of a helical dislocation is one of constant pitch and loop radius [13]. Since a nonequilibrium shape is sometimes observed, we conclude that in the sample treated at the lowest scan velocity the annealing time is still not long enough for all the dislocations to reach this equilibrium situation. This in turn can be caused by the continuous formation of new helical segments, which means that not all vacancies are used, or by an increased locked segment through relaxation to the equilibrium shape. These explanations are also in line with the fact that the hardness curve is still rising at the $\frac{1}{8}\,\mathrm{cm\,s^{-1}}$ point (see Fig. 18). Nevertheless, we do not expect that this mechanism will dominate over the role of homogenization as described previously.

High Scan Velocity Regime

The increased hardness in the high-velocity regime, reflects the generally observed trend, and is caused by two reasons. First, the cell size decreases upon increasing the laser scan velocity, and causes a hardening that depends on K and a in the following equation: $\tau = K/d^{\alpha}$. Unfortunately, the values of α and K are not easily derived. α may vary from $\frac{1}{2}$, in the Hall–Petch relation, up to 2, as has been described by Hegge and De Hosson [56]. Furthermore, precipitates will contribute to the hardness, and this contribution may be even larger than the contribution arising from the small cell size. The presence of these precipitates is observed in samples with scan velocities of 1 to $4\,\mathrm{cm\,s^{-1}}$.

To calculate the contribution of the precipitates to the hardness one needs to know the average size and volume fraction as deduced from TEM observations. Since the θ' precipitate is commonly regarded as giving the most important contribution, attention will be paid only to this precipitate.

Because the shape of the θ' precipitate deviates strongly from spherical, some corrections to the observed dimensions are required to obtain the dimensions in the glide plane. In the glide plane, the mean length of the obstacles D' is $\pi D/4$, and the mean thickness d' is $d\sqrt{3/2}$. Given a volume fraction f_v, the mean center-to-center spacing is: $\lambda' = \sqrt{D'd'/f_v}$.

A distribution of linear parallel obstacles of length $L = D' - d'$ will contribute to an increase in flow stress that is $1 + (L/\lambda')$ times greater than for the associated distribution of point defects. The latter applies to aligned obstacles only. If we take into account the finite thickness of plate-shaped particles, the effective separation between the precipitates is [73]

$$\Lambda_p = \frac{\lambda'}{1 + L/\lambda'} - d' \tag{16}$$

The flow stress is then given by

$$\tau = \tau_0 + \frac{k_1 k_2}{2\pi \Lambda_p} Gb \ln\left(\frac{\Lambda_p}{b}\right) \qquad (17)$$

where k_1 is a statistical factor equal to 0.85, and k_2 is a factor depending on the character of the dislocation varying between 1 and 1.5 [74]. The contribution to the flow stress arising from other effects is represented by τ_0. From TEM observations, $D = 150\,\text{nm}$, $d = 6\,\text{nm}$ and $f_v = 0.02$. Substituting these values into Eqs. (16) and (17) results in an Orowan stress of 49 MPa.

The absence of GP zones in samples that have not yet developed precipitates, in the range between 8 and $25\,\text{cm s}^{-1}$, might be explained by the fact that the amount of copper in solid solution is less compared to conventionally heat-treated Al 2024. In aluminum with 2 wt.% of copper, GP-I zones do not appear, and the precipitation sequence starts with GP-II [69]. Furthermore, it has been reported that in rapidly quenched samples the early stages in the precipitation sequence are accelerated and that GP-II zone formation is suppressed in favor of the θ' phase. If both effects are present in our samples, the θ' phase is indeed the first precipitate to be observed.

An additional point that might explain the absence of the GP zones is that these precipitates are related to the stress condition of the matrix. Since GP zone formation is accompanied by a decrease in average atomic volume, the resultant hydrostatic tensile stresses will contribute to the formation energy. When a hydrostatic tensile stress is already present, the formation of GP zones will be hindered. In our samples, tensile stresses arise because of shrinkage during solidification. These stresses are of the order of 100 MPa, which is of the same order of magnitude as in Al–Cu containing silicon or Al_2O_3 particles. As a consequence, GP zone formation may also be hindered by the presence of the hydrostatic stresses in our case.

When explaining the hardness curve versus laser scan velocity (see Fig. 18) based on the above-mentioned TEM observations, it will become clear that the contributions of cell boundary strengthening and solid solution strengthening do not completely explain the decrease in hardness with increasing laser scan velocity in the lower scan velocity regime. In fact the contribution of the cell boundary strengthening increases over the whole regime. Taking the proportionality constant K of the Hall–Petch relation for cell boundaries (upper limit) to be one-half of that for grain boundaries (about $0.07\,\text{MN m}^{-3/2}$ [75]) the cell boundary strengthening may increase from 20 to 35 MPa in the low-velocity regime, whereas the contribution by solid solution strengthening decreases from a maximum of 19 MPa, based on 2.32 wt.% of copper in solid solution in the samples treated at the lowest laser scan velocity of $\frac{1}{8}\,\text{cm s}^{-1}$. In total it results in a much too small increase in hardness at the lowest scan velocity of $\frac{1}{8}\,\text{cm s}^{-1}$. So, in this regime the immobilizations of dislocations by the helical segments and by preferential nucleation of solutes have also to be considered to explain the increase in the flow stress in a more quantitiative way.

After shot peening, an increase in hardness independent of the laser scan velocity is observed. This is not what one would expect since the microstructure is quite different at different scan velocities.

However, the precipitation induced by shot peening that was observed earlier in a laser-melted Al–Si alloy explains why the observed values are the same for the low and high scan velocities: after shot peening, precipitates are present independent of the scan velocity regimes, whereas the amount of alloying elements in solid solution is expected to be much less than before shot peening.

A calculation applying the same methods as in the samples with platelets before shot peening results in a contribution to the flow stress of 76 MPa. The TEM values taken for the size and volume fraction are $D = 100\,\text{nm}$, $d = 5\,\text{nm}$ and $f_v = 0.03$. The 3% volume fraction is also in agreement with a comparison with the 2.32 wt.% copper alloy in which the cell walls are just fully dissolved in $\frac{1}{8}\,\text{cm s}^{-1}$ tracks, thereby taking into account that copper is also participating in the S′ precipitates in Al 2024.

Obviously after shot peening the contribution of helical dislocations to the change in hardness is not relevant, whereas the contribution from the increased dislocation density is approximately the same in all cases. Following the same methods as before, the contribution to the flow stress amounts to 25 and 79 MPa for dislocation densities of 2.5×10^{15} and $2.5 \times 10^{16}\,\text{m}^{-2}$, respectively.

The minimum value of the hardness of samples treated at a scan velocity of $\frac{1}{2}\,\text{cm s}^{-1}$ is still present after shot peening, and is also visible in the curves of the residual stresses as well as of the dislocation density versus laser scan velocity. Apparently the amount of alloying elements brought into solid solution during laser melting has a minimum for this scan velocity. At higher scan velocities, more copper and magnesium is retained during the rapid solidification process, and for the lower ones the homogenization time during cooling is longer.

The remarkable feature that the hardness curve observed has a minimum, at a laser scan velocity of $\frac{1}{2}\,\text{cm s}^{-1}$, is quite different from the monotonically increasing curves, usually detected in many other laser-treated materials. The reason for this observation can be explained by TEM investigations, and is linked to the minimum in the solid solution concentration. By varying the quench rate it is possible to achieve different stages in the precipitation sequence: it turned out that depending on the laser scan velocity, nucleation and growth of precipitates could either be enhanced or suppressed. The formation is suppressed by quenching too rapidly using a high laser scan velocity, whereas at too low a scan velocity the formation is suppressed due to a lack of sufficient vacancies. In the latter case, the vacancies are trapped by dislocations, leading to the formation of helical dislocations, that is, before the temperature is low enough to develop a driving force for precipitation.

The mechanisms by which the hardening occurs are the same as the ones in conventionally hardened specimens, except for the contributions originating from the small cell sizes and from the formation of helical dislocations. The latter, however, do not seem to contribute significantly. Only at lower laser scan velocities does dissolution of cell walls lead to an increased amount of copper and magnesium in solid solution and, therefore, to an increase of the hardness caused by solid solution hardening.

From comparison of the absolute hardness values between the laser-treated samples and the T3 samples, one might conclude that laser surface melting is not a suitable technique for obtaining hard surfaces. This conclusion is wrong, since the amount of copper that is brought into solid solution is far less than what ultimately is possible. In fact, one should not use a hypoeutectic alloy when a high solid solution concentration is required. In the past it has been shown, by applying splat cooling of a eutectic Al–Cu alloy, that all of the copper (33%) can be retained in solid solution [76]. Furthermore, it was shown that the precipitation sequence was still the same as in the more dilute alloys. A higher concentration of copper obtained by laser melting of such an alloy will likely lead to a high density of θ' platelets, resulting in substantially higher hardness values.

Cracking of the samples limits the practical applicability of laser treatments of this alloy. In order to prevent cracking, it might be beneficial to use an alloy with a high copper content, with possibly some supplementary silicon, since these elements reduce the thermal expansion coefficient. A higher copper content thus results in a higher strength and a reduced thermal expansion.

So far we have restricted ourselves to the derivation of the physical properties of metals from static dislocation configurations. Of course the collective dynamic behavior of groups of dislocations determine these properties. Therefore, dynamic TEM observations are extremely relevant but sometimes rather complicated to analyze thoroughly. This will be the topic of the next section.

1.4.2 Dynamic Transmission Electron Microscopy Observations: In Situ Deformation

As far as metallic systems are concerned, a vast number of TEM investigations have been devoted to intermetallic systems. Intermetallics represent a class of materials with interesting mechanical properties for applications. Why? Most metals having an f.c.c. structure are ductile and their strength decreases with increasing temperature. In contrast, intermetallic ordered compounds having the $L1_2$ structure (space group $Pm3m$) show many remarkable and unusual mechanical properties. For example, many exhibit a temperature anomaly: a rise of their strength with increasing temperature, up to many hundreds of degrees. Besides the strength at high temperatures, the moderate strength at lower temperatures has the additional advantage that shaping at room temperature is relatively easy. These properties, together with the chemical resistivity that most ordered compounds show, make this class of materials very versatile for applications in hostile environments. In general these materials are very attractive for high-temperature applications. However, although many of the ordered compounds with strong ordering tendency are ductile as single crystals, they show brittle fracture in their polycrystalline form. As a matter of course this hampers considerably their application.

Ni_3Al, with an $L1_2$ structure, is one of the ordered compounds that is a very promising candidate for various applications, such as turbine blades used by the aircraft industry. Aoki and Izumi [77] showed that the ductility at room

temperature of nickel-rich polycrystalline Ni_3Al can be improved notably by addition of small amounts of boron. However, the increase in ductility drops again at high temperatures. If useful structural materials based on these intermetallic ordered compounds are to be further developed it is crucial to scrutinize the reasons for their brittle behavior. A possible reason could be low cohesion of the grain boundaries. If the energy of a grain boundary is not much lower than the sum of the energies of the two surfaces that are created by fracturing along the grain boundaries, intergranular fracture can occur by simple de-cohesion of the grain boundary.

The work presented here, however, deals with a different approach by considering explicitly the plasticity near grain boundaries, as deduced from TEM observations. Consequently, the ultimate objective is to establish a link between intergranular brittleness and the interaction between dislocations and grain boundaries. In this way, the macroscopic mechanical properties of a number of ordered compounds are explained in terms of the atomic structure that is present in these materials.

There exists experimental evidence that in Ni_3Al the dislocation mobility in the vicinity of grain boundaries may be strongly enhanced when ductilization takes place [78] and that plastic flow precedes intergranular fracture [79]. Considering these experiments, it might be reasoned that the passage of gliding dislocations arriving from the lattice might be hindered by grain boundaries. A third point of view is that in ordered compounds, the grain boundary itself might have less ability to act as a source of dislocations. Here, in situ TEM deformation studies of Ni_3Al are presented. In this type of experiment, the development of the interaction between dislocations and grain boundaries can be followed and characterization of these defects can be carried out.

Various approaches exist to the experimental study of the interaction between lattice dislocations and grain boundaries. Slip line analysis [80] reveals the steps in the surface of a deformed sample that dislocations leave behind as they glide. There are only steps in the surface if the Burgers vectors of the dislocations are not parallel to the surface. The surface of the sample should be polished before deformation, to be able to observe these surface steps, called slip lines, with an optical microscope. Etch pitting [81] displays the site where the dislocation emerges at the surface. If a surface is etched, atoms are preferentially removed at the position of the dislocation line, and in this way an etch pit is created. Both techniques can show the changes in the propagation of dislocations when a slip band containing many dislocations impinges on a grain boundary. However, these techniques are not capable of giving information about the nature of the dislocation and the grain boundary, such as the line direction and the Burgers vector of a dislocation, the orientation of the grain boundary plane and the misorientation between the two grains. In contrast, TEM is a technique that allows observation of defect configurations in thin foils and also provides information on the nature of the defects.

In many TEM experiments (e.g., see Dingley and Pond [82] and Elkajbaji and Thibault-Desseaux [83]), bulk samples of different materials have been deformed, and these samples have been prepared for

study in the transmission electron microscope. Although a full analysis of the dislocation–grain boundary configuration is possible, the development of the interaction has to be deduced from the configuration that is left behind after the interaction has taken place.

In situ deformation in a transmission electron microscope is one of the very few techniques by which the development of the interaction between lattice dislocations and grain boundaries can be studied and which at the same time allows analysis of the configuration. By this technique, samples which have not yet been deformed are strained inside the microscope in a special straining holder. In practice there are a number of complexities. The dislocations may have high velocities and therefore they may be difficult to observe when moving. As the strains are usually very small in in situ experiments, the deformation is not homogeneous and therefore it may take place in a different part of the sample: in the region that is not transparent for electrons or in a part of the electron transparent region that is not in contrast at that moment. In addition, the fact that the interaction is being studied in a very thin foil could have a substantial influence (remember that we are interested in bulk properties): the electron transparent region may have a complicated geometry because of the thinning and therefore the stresses in the thin region may be different from the bulk stress state. Further, an oxide layer on the surface may hinder the motion of dislocations. Therefore, care must be taken to study the interactions in the thicker regions of the foil. Nevertheless, in many cases the in situ TEM technique is a promising tool for investigations of dislocation–grain boundary interactions. As the state prior to deformation is known, the changes that have been introduced because of the deformation can be deduced even if the interaction has not been observed directly. Analogous to the slip lines at the surface of a bulk sample, each dislocation that moves may leave behind an atomic step in the surface of the thin foil. Under the right imaging conditions, these atomic steps give rise to diffraction contrast and, therefore, even if a dislocation has not actually been seen to move, it may be possible to deduce the motion of a dislocation by the slip traces that it has left.

To transfer slip from one grain to another in a polycrystalline material, various mechanisms may occur which depend on the misorientation between the grains and the grain boundary plane itself. For those readers not familiar with some of the concepts in the description of grain boundary structures a very concise summary is presented on the coincidence site lattice (CSL) and the displacement shift complete (DSC) lattice which are used in the TEM analysis of the interaction between lattice dislocations and grain boundaries in ordered alloys.

1.4.2.1 Grain Boundary Structures

A grain boundary interface has eight degrees of freedom: the misorientation axis and angle, the boundary plane normal and the relative translation of the two grains with respect to each other. For ordered alloys the position of the boundary plane in the direction of its normal, corresponding to different chemical compositions of the boundary plane, is an

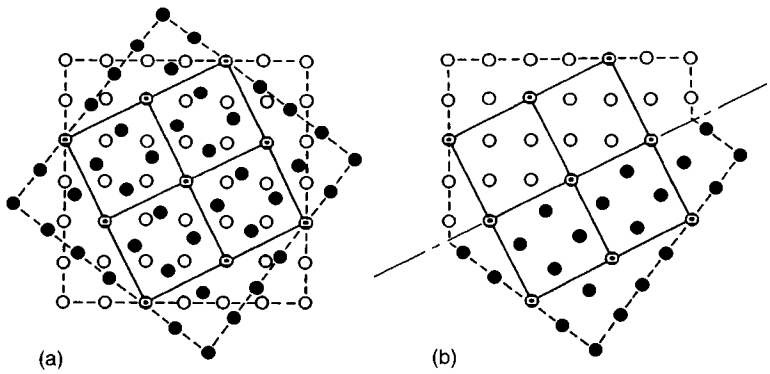


Figure 24. (a) The dichromatic pattern of a $\Sigma = 5$ boundary in a simple cubic structure. Projection along the [100] direction. Coincidence of black and white sites is indicated by black sites containing a white circle. The solid lines connecting the coincident sites indicate the CSL. The CSL periodicity in the [100] direction out of the plane of the paper is equal to the lattice periodicity. (b) Grain boundary constructed on the basis of the dichromatic pattern of (a). The broken line indicates the grain boundary plane.

extra degree of freedom. A model that is not restricted to low misorientation angles is the CSL model [84, 85].

This model uses the dichromatic pattern which is created by hypothetically allowing the lattices of the two grains to interpenetrate. The lattice sites belonging to grain 1 are considered white whereas the lattice sites of grain 2 are black. For certain misorientations a new (super-) lattice of coincident lattice sites of the white and the black lattice, a CSL, exists (Fig. 24). The parameter Σ is defined as the ratio between the volume of the CSL unit cell and that of the primitive unit cell. The density of coincident lattice sites in space equals $1/\Sigma$. A grain boundary is constructed by inserting a plane in the dichromatic pattern and letting the atoms occupy the white sites on one side of the plane and the black sites on the other side of the plane. The location of the plane can be chosen such that it intersects coincidence points (Fig. 24). In the early stages of the development of this model, it was reasoned that a high density of coincidence sites in the boundary plane would mean a low grain boundary energy. Nowadays, it is considered that this relation is not so simple, but for some low boundaries (and thus for a high density of coincidence sites) lower energies are indeed found. Also, it has been established that there may exist a relative translation away from exact coincidence for the equilibrium structure of a low Σ boundary, and the individual atoms at the grain boundary plane may have been displaced. It has to be noted that the CSL is a mathematical model: an infinitesimal rotation or translation of one grain with respect to the other destroys the whole CSL. Nevertheless, a grain boundary with the same misorientation but translated away from coincidence is still regarded as being a boundary of the same special coincidence and as having the same Σ value. The translational symmetry of such a boundary is the same as that of the boundary in exact coincidence. Grain boundaries in the ordered structure may have different ordering configurations, that is, there exist configurations that are topologically identical, but the positions of the A and B atoms are different [86]. The

ordering configuration with the lowest energy can be regarded as the equilibrium structure of the grain boundary. The DSC lattice [85, 87] of a bicrystal in CSL orientation is the lattice of the displacement vectors of one grain with respect to the other that keep the dichromatic pattern (black and white sites) unchanged except for a translation. In other words, if the black lattice is displaced over a vector belonging to the DSC lattice, the black atoms that were in coincidence with the white atoms may have lost their coincidence, but somewhere else black and white atoms will now be in coincidence and the dichromatic pattern (and the CSL) will have the same shape as before (Fig. 25). The DSC lattice is the coarsest lattice that contains both crystal lattices of the two grains as sublattices. It has smaller unit vectors than the unit vectors of a perfect lattice. Note that each CSL has a specific DSC lattice associated with it.

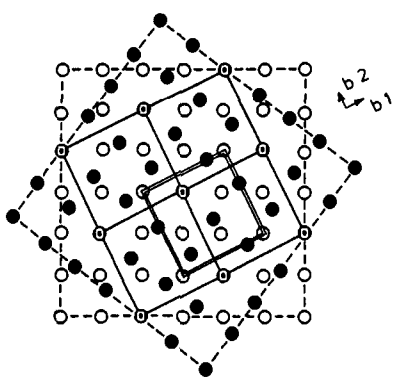


Figure 25. Same boundary as Fig. 24. The basis vectors $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ of the DSC lattice belonging to this boundary are indicated. $\boldsymbol{b}_3$ points out of the plane of the paper. The double lines indicate the new location of a CSL unit cell if the black lattice is translated with respect to the white lattice over $\boldsymbol{b}_1$. The vector connecting the original and the new position of the CSL unit cell is called the step vector associated with the $\boldsymbol{b}_1$ DSC vector.

The fact that the dichromatic pattern is unchanged for a displacement over a DSC vector of one grain with respect to the other has consequences for the types of Burgers vectors that are allowed for dislocations in the grain boundary plane. A perfect dislocation must have a Burgers vector that is a translation vector of the material in which it exists. Now, if we have a grain boundary, we have just seen above that a displacement over a vector that belongs to the DSC lattice leaves the dichromatic pattern unchanged except for a translation. Thus, in a grain boundary, grain boundary dislocations (also called DSC dislocations) with Burgers vectors belonging to the DSC lattice can exist. The displacement over a DSC vector may cause a shift in the dichromatic pattern, over a vector, called the step vector (Fig. 25).

The dislocation is then associated with a step in the grain boundary plane, located at the core of the dislocation [88]. The height of the step in the grain boundary plane can conveniently be expressed in terms of interplanar spacings, that is, the distance between crystallographic planes parallel to the boundary plane. Dislocations in a grain boundary having a Burgers vector that belongs to the DSC lattice and that is not a lattice translation vector of either grain can only exist in the grain boundary, and the movement of these dislocations is reconstricted to the grain boundary plane. As the DSC lattice is specific to a CSL orientation, each grain boundary has specific allowed Burgers vectors for grain boundary dislocations.

A number of interaction processes between lattice dislocations (the designation 'lattice' is used to discriminate these from grain boundary dislocations) and

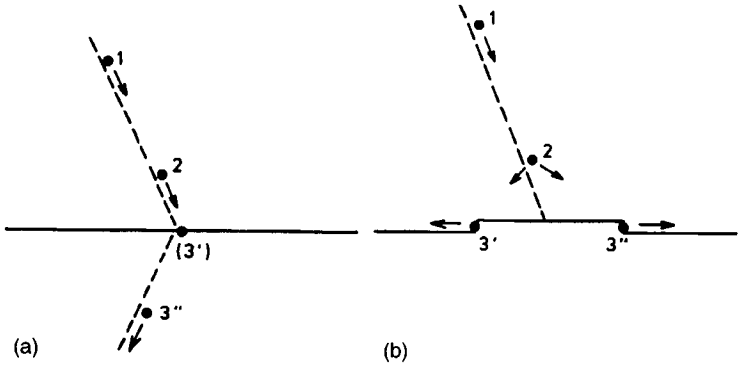


Figure 26. Interaction mechanisms between lattice dislocations and grain boundaries. (a) Transmission: another lattice dislocation emerges at the other side of the grain boundary and a residual dislocation may be left at the grain boundary. (b) Absorption: the lattice dislocation dissociates in the grain boundary plane into DSC (or grain boundary) dislocations, and a step in the grain boundary plane is created between the DSC dislocations.

grain boundaries may be discriminated (Fig. 26). A dislocation may move into the grain boundary, while another lattice dislocation emerges from the grain boundary into the other grain. As the other grain has a different orientation, the outgoing dislocation may have a Burgers vector that differs from the incoming dislocation. A residual dislocation remains in the boundary plane with a Burgers vector that is equal to the difference in the Burgers vectors of the incoming and outgoing dislocations. The residual dislocation is always a DSC dislocation. This mechanism is referred to as transmission. Another possibility is that the lattice dislocation can dissociate into grain boundary dislocations or DSC dislocations. In this way the elastic energy of the dislocation can be reduced. This mechanism is referred to as absorption. Between the grain boundary dislocations, a step in the grain boundary plane may be created. The atoms in the region which belonged to the lower grain before the lattice dislocation was absorbed and after absorption belong to the upper grain have to adapt their positions to the lattice of the upper grain. If the step in the grain boundary plane is larger than one interplanar spacing, the adaptation cannot be achieved by means of a simple rigid-body translation, but a less ordered relocation, called shuffling, of the atoms will take place. Analogous to lattice dislocations, grain boundary dislocations may dissociate in the grain boundary plane into partial grain boundary dislocations, creating a fault in the grain boundary plane between them. In addition there are more complex mechanisms, such as absorption and reemission of a lattice dislocation at another site in the boundary.

The grain boundary dislocations that stem from an absorbed lattice dislocation or that are the residual dislocations after transmission of a lattice dislocation are called extrinsic grain boundary dislocations, as opposed to intrinsic grain boundary dislocations, which form the secondary dislocation network in a grain boundary. Besides these mechanisms describing the

interaction between lattice dislocations and grain boundaries, dislocations may be nucleated at the grain boundary. A grain boundary can act as a source of dislocations that can be generated under the influence of a stress field. The dislocations can be generated in either grain and in the boundary plane.

In a CSL model of the grain boundary structure, moving dislocations which are of course lattice dislocations and therefore also vectors of the DSC lattice may interact with the boundary structure in various ways, for example the dislocation forming a line of intersection at the boundary may lower its energy by the generation of secondary dislocations with smaller DSC vectors [89–91]. It depends on the orientation of these new secondary dislocations how they react with the already existing secondary dislocations of the boundary structure. To the best of our knowledge, Schober and Balluffi were the first to observe such interactions in (001) twist boundaries [92], followed by others extending these observations to a whole range of twist as well as tilt boundaries [93–99]. For example it was observed that the unitary $\frac{1}{2}\langle 110\rangle$ lattice dislocations in polycrystalline steel dissociated into five secondary grain boundary dislocations upon interaction with a $\Sigma 29$ grain boundary [92, 94]. Indeed, although the Burgers vector were not analyzed in detail, this observation is consistent with the dissociation reaction into two sets of partials, $2(\frac{1}{58})[10\,4\,0]$ and $3(\frac{1}{58})[370]$, which are DSC vectors of $\Sigma 29$.

However, despite these observations on the interaction between lattice dislocations and secondary grain boundary dislocations, there has been quite a debate in the literature on the necessary and sufficient conditions under which dislocations transmit through the grain boundary. This has to do with the question what slip system might be activated in the other grain. Several of the criteria are summarized by Shen et al. [100, 101]:

(i) The following summation should be a maximum for the active slip system in the other grain:

$$(\boldsymbol{\eta}_1 \cdot \boldsymbol{b}_2) \times (\boldsymbol{\eta}_2 \cdot \boldsymbol{b}_1) + (\boldsymbol{\eta}_1 \cdot \boldsymbol{\eta}_2) \times (\boldsymbol{b}_1 \cdot \boldsymbol{b}_2),$$

where $\boldsymbol{\eta}$ are the slip plane normals and $\boldsymbol{b}_i$ are the slip directions in both crystals.

(ii) The force on the transmitted dislocation is a maximum for the active slip system in crystal 2.

(iii) Another conceptual idea is that the dislocation line has to rotate from the line of intersection between the boundary plane and the slip plane in crystal 1, $\boldsymbol{\xi}_1$, to the corresponding line of intersection in crystal 2, $\boldsymbol{\xi}_2$, before transmission may occur. Consequently the criterion is to maximize $(\boldsymbol{\xi}_1 \cdot \boldsymbol{\xi}_2)(\boldsymbol{b}_1 \cdot \boldsymbol{b}_2)$.

(iv) Finally, one should derive the slip direction from (ii) and the slip plane from (iii).

Although Shen et al. [100, 101] were able to predict the correct slip direction and active slip plane using these criteria, it was shown by Lee et al. [102] that the magnitude of the residual grain boundary dislocation should also be considered explicitly. They state:

(i) The outgoing slip plane is determined by minimizing the angle between the intersections of incoming and outgoing slip planes, where $\boldsymbol{\eta}_1 \cdot \boldsymbol{\eta}_2$ should be maximized.

(ii) The magnitude of the Burgers vector of the residual dislocation left in the grain boundary should be a minimum.
(iii) The outgoing slip system should have the maximum resolved shear stress from the piled-up dislocations.

Criteria (ii) and (iii) may be contradictory with respect to the prediction of the Burgers vector of the outgoing dislocations. Lee et al. [102] find in their observations that criterion (ii) seems to dominate.

1.4.2.2 TEM Holder for Straining at Temperature

Ni_3Al was prepared by arc melting 99.99% pure nickel and 99.999% pure aluminum. The material was homogenized at 1100°C (1373 K) for 5 days, resulting in grain sizes of 1 mm. Miniature tensile specimens (6×3 mm, with a thickness of 350 μm) were cut out of the bulk material by spark erosion. There were two holes in the sample through which it was held by the pins of the two grips of the deformation holder (see also Fig. 39).

The sample was necked in the middle, so as to maximize the likelihood that the deformation would start near the future electron transparent region. Care was taken to have a grain boundary, preferably a coherent twin boundary, present in the middle of the sample.

Next, the sample had to be thinned, to obtain an electron transparent region. First, the sample was ground to a thickness of 150 μm using a Gatan disc grinder, and then the sample was dimpled on both sides at the location of the grain boundary, reducing the thickness locally from 150 μm to around 60 μm. This was done to increase the probability of having the grain boundary in the thin area. The last step of the preparation process was the final thinning of the specimen in a Struers Tenupol electropolishing unit, using a mixture of 70% methanol and 30% nitric acid at 0°C, at an applied voltage of 8 V. In the Tenupol apparatus, the sample is thinned by two jets of electropolishing fluid, which are directed at both sides of the specimen. Platinum diaphragms having a hole with a radius of 2 mm were used to ensure that only the center of the sample could be reached by the electropolishing fluid. If the right voltage is applied between the fluid and the sample, material will be removed from both sides of the specimen by the electropolishing process. The polishing process was stopped when enough material had been removed from both sides such that the sample was perforated in the center. The perforation is detected by light shining through the sample, and immediately after perforation the process stops automatically. The area around the hole now is transparent for electrons. The specimen was mounted in a special single-tilt in situ deformation holder, built on principles taken from Kubin and Veyssière [103]. The interesting point here is that in situ deformation experiments can be carried out also at high temperature. Commercially available holders usually separate elongation from high-temperature applications. We made the tip of the holder, 95 mm long, of the nickel-base superalloy Nimonic 115, which exhibits excellent creep properties. A temperature of 1000°C can be achieved by dissipation of about 13 W in a tungsten filament that heats the specimen by radiation rather

than by conduction. Temperature measurements were done via a thermocouple, and minimization of heat losses (mainly by radiation) was achieved by enclosing the specimen in a double box of gold sheet. The boxes (except the bottom) as well as the filament can be taken out for mounting the specimen within the limited space available. The filament can be taken out of the holder easily, because the electrical contacts of the filament with the wires to the power supply consist of flexible platinum-sheet lips. The specimen itself is held by two Nimonic pins, a fixed one on the tip side of the holder and a mobile one which is connected by a pulling rod with metal bellows on the outside end of the holder, providing a vacuum seal, and a second metal bellows, connected to the primary vacuum pump of the microscope. A maximum load of about 30 N can be applied. An electrical control system offers the possibility of cyclic straining, and the strain rate can also be varied. The pressure in the bellows is measured by a transducer (1 mbar to 1 bar $\pm$ 1%) with a recorder. The control system consists of four valves. Two solenoid-operated isolation valves determine whether the bellows shrink by pumping or expand by air inlet. Two needle valves determine the rate of expansion or shrinkage. The strain rate as well as the minimum and maximum pressure can be chosen at the control panel. In situ creep experiments can be done with the same equipment by reducing the amplitude of the pressure cycle to zero.

The advantage of the pneumatic straining is the absence of vibration, which offers the possibility of TEM observations of crack growth or dislocation motion in front of a crack tip during cyclic straining. For example, it has been observed that a triangular variation in pressure (amplitude 25 mbar) around 720 mbar causes cyclic straining with an amplitude of about 100 nm. A typical value for the cycle time is 30 s, and therefore strain rates $\dot{\epsilon}$ are in the domain of conventional values (10^{-4}–$10^{-5}\,s^{-1}$). Although tilting is limited to a single tilt ($\pm 30°$), one or two suitable diffraction conditions (two beams, $s_g \neq 0$) can be obtained generally. The processes during deformation could be monitored by a TV system and recorded on video. The deformation was stopped before total disruption of the specimen.

The deformation holder has only one tilt axis. Therefore, the specimen usually cannot be tilted to the right with respect to the electron beam so as to excite the different reflections that are used in the analysis of the defects present in the sample. As a consequence, after the in situ deformation, specimens that promised to be interesting were shaped into 3 mm diameter disks by very cautious grinding, taking care that the thin area was protected. In this way they fitted into a TEM holder with two tilt axes, and the defects could be analyzed in detail.

The determination of the orientation relationship between the two grains was done as follows. The crystallographic direction of the incident electron beam can be determined using the diffraction pattern of a crystal. When the Kikuchi lines are visible, the incident beam direction can be determined very accurately. Kikuchi lines are formed by inelastically scattered electrons that are subsequently rediffracted. These lines indicate the orientation of the crystallographic planes by which the electrons are rediffracted. The direction of the incident beam is determined by measuring the distance of

several Kikuchi lines to the undiffracted beam in the diffraction pattern. The incident beam direction is determined in each grain, taking care to keep the orientation of the bicrystal fixed while moving from one grain to another. In this way, two parallel crystallographic directions in the two grains are found. By determining the incident beam direction in the two grains several times, with different orientations of the bicrystal, a rotation matrix transforming the coordinate system of one grain into that of the other grain is determined. As the crystals have cubic symmetry, this rotation matrix is not unique. Using a computer program the equivalent rotation matrices describing the same orientation relation are also calculated. From the set of equivalent matrices, the characteristic misorientation can be determined by finding the matrix describing a 180° rotation. All low boundaries have a rotation axis that belongs to a 180° rotation. The Miller indices of this axis give the characteristic misorientation [104]. For example, the twin boundary, $\Sigma = 3$, has its 180° rotation axis along [111].

1.4.2.3 In-Situ Deformation Results

Dislocation-Grain Boundary Interactions in Ordered Structures

In many samples, cracks (some of which were already present before the deformation started) were observed that had initiated at the edge of the thin foil, propagating along planes. Dislocations were often seen to be emitted from the crack tip, in the plane of the crack and sometimes also on inclined planes. In addition, observations were made of dislocations arriving from the bulk, although these observations were not so numerous. The dislocations from the bulk never arrived all on the same slip plane, but seemed to appear in slip bands. Quite frequently, the propagation of the crack occurred with a jerky type of motion, and it was then impossible to observe any dislocation motion.

In many of the specimens there was a grain boundary, often a coherent twin boundary ($\Sigma = 3$, 109.47° around [110] with a boundary plane), visible in the electron transparent region. A number of observations were made of individual dislocations that had been emitted from cracks. Sometimes dislocations arriving from the bulk impinged on twin boundaries and were arrested at the boundary plane. Cracks were seen that had grown through a twin boundary and had changed their direction of propagation upon crossing of the boundary plane.

One sample showed a crack which had grown during the in situ deformation to the close vicinity of a coherent twin boundary, but which had not crossed the boundary (Fig. 27).

On the other side of the boundary, starting exactly from the line of intersection of the crack plane and the boundary plane, slip traces could be observed leading into the other grain, to a large number of dislocations that all had the same slip plane (Fig. 28).

This sample was chosen for further analysis in the double-tilt holder. The rotation of the boundary under study could be described within the margins of error as a 109.5° rotation around [110], characteristic for a twin boundary. By tilting to an edge-on position, the

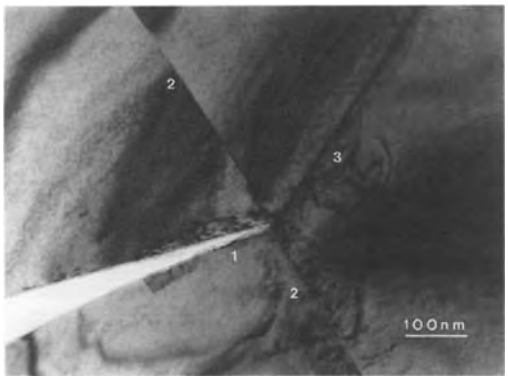


Figure 27. A crack that has grown very close to a $\Sigma = 3$ coherent twin boundary and slip lines on the other side of the boundary. 1, crack; 2, boundary plane; 3, slip lines.

boundary plane was determined to be $(\bar{1}11)$, which is equal to $(\bar{1}1\bar{1})_{II}$. The index 'II' indicates the coordinate system of the grain containing the dislocations. The grain containing the crack is denoted if no index is used. In a similar way, the plane of the crack was determined to be close to $(1\bar{1}1)$ and the slip plane of the dislocations was determined to be $(1\bar{1}\bar{1})_{II}$. By the $\boldsymbol{g} \cdot \boldsymbol{b} = 0$ invisibility criterion, the Burgers vector of the dislocations was determined to be parallel to $[110]_{II}$; this is the [110] direction that is common to both grains. The line direction was determined to be $[230]_{II} \pm 13°$, which is close to the $[110]_{II}$ screw direction.

Let us concentrate on the interaction between lattice dislocations with low-angle grain boundaries. In one sample, in the thin electron transparent region many low-angle grain boundaries were present, but no high-angle grain boundaries were visible. In this sample there were only a few holes of approximately 5 μm in diameter, and there were no cracks in the sample at all, in contrast to most samples that contained a hole of about 50 μm in diameter

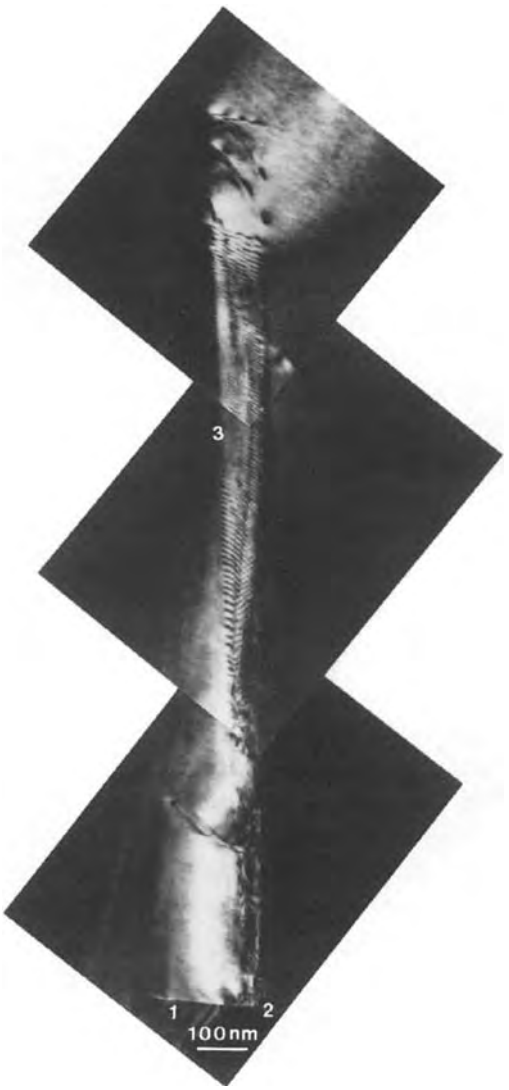


Figure 28. The dislocation configuration that was found attached to the slip lines. 1, crack; 2, boundary plane; 3, dislocation array.

and a few small initial cracks. The external force that had to be applied to cause deformation in the thin region was twice as high as usual for samples of the same thickness. This might be explained by the absence of stress concentrations because of cracks or large holes. During the

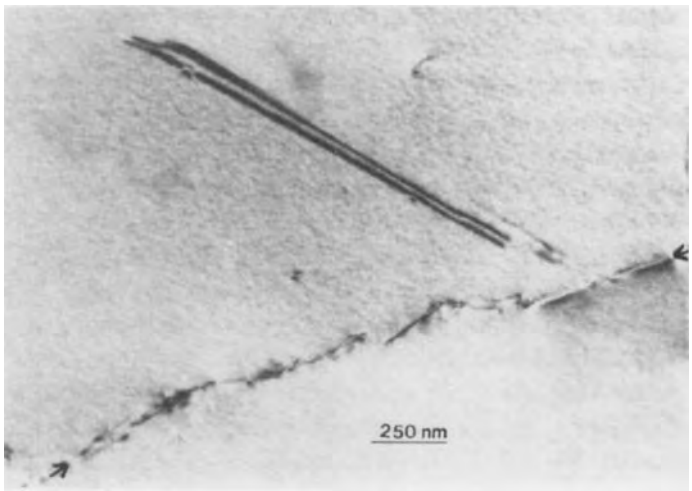


Figure 29. Stacking fault near a low-angle grain boundary (arrows), showing characteristic fringe contrast.

experiment, many dislocations were seen gliding, almost exclusively in one slip band, in parallel planes. It was observed that at several places elongated stacking faults, hundreds of nanometers long, had been created at the intersections of the slip band with low-angle grain boundaries (Fig. 29). The faults always lay at the same side of the boundaries: the direction of motion of the gliding dislocations always pointed from the boundary to the faults. Thus, it can be assumed that the faults have been created by interaction of the gliding dislocations with the dislocations in the boundary. Also, at several other locations along the slip band, away from the intersections with low-angle grain boundaries, gliding dislocations had formed faults. The deformation experiment was stopped after a crack had developed, originating from one of the holes. The sample was shaped into a 3 mm disk for further investigation in a double-tilt holder. One of the intersections of the slip band with a low-angle grain boundary where faults had formed was chosen for detailed analysis. Three faults can be seen that are close to three dislocations in the low-angle grain boundary, at distances varying from less than 50 to 200 nm (Fig. 30).

The foil normal was $[10\,3\,11] \pm 3°$, which is 11° from [101], and the foil thickness at the point of interest was 110 ± 15 nm. If we assume that the macroscopic tensile axis was perpendicular to the foil normal and parallel to the line

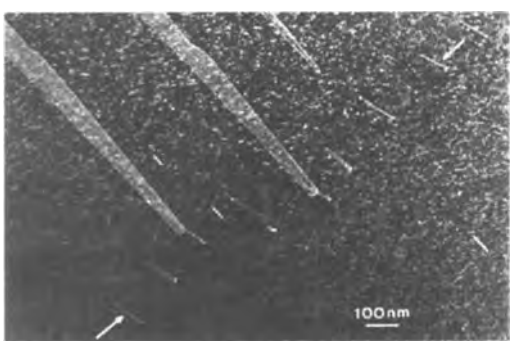


Figure 30. Weak-beam micrograph of superlattice intrinsic stacking faults (SISFs) near a low-angle grain boundary. The arrows indicate the location of the low-angle grain boundary or cell wall, consisting of an array of parallel dislocations spaced about 200 nm apart. Only one of the three SISFs can be seen fully; the other two are only partially in view.

connecting the two grips by which the sample was held, this axis could be determined to be $[4\bar{7}\bar{2}] \pm 10°$. It has to be noted that similar fault configurations close to a low-angle grain boundary were observed in thicker parts of the specimen, farther down the same slip band. The average line direction of the dislocations constituting the low-angle grain boundary was $[02\bar{5}] \pm 5°$. The dislocations in the boundary had two different Burgers vectors: $[1\bar{1}0]$ or $[110]$. The dislocations that were close to the faults all had a Burgers vector of $[1\bar{1}0]$. The dislocations in the boundary that had $\boldsymbol{b} = [110]$ were dissociated so widely into two $\frac{1}{2}[110]$ superpartial dislocations in the (100) plane, that the individual partials could be resolved by weak beam microscopy (Fig. 30).

The plane of the faults was determined to be $(1\bar{1}1)$ by tilting the faults to an edge-on position. The fault vector $\boldsymbol{R}_f$ was determined to be $\pm\frac{1}{3}[1\bar{1}1]$ by applying the invisibility criterion $\alpha = 2\pi \boldsymbol{g} \cdot \boldsymbol{R}_f = n2\pi$. However, for a fault of this type the positive and the negative fault vectors are not equivalent. In one case, a $(1\bar{1}1)$ plane is removed (superlattice intrinsic stacking fault, SISF; this is a shear fault) and in the other case an extra $(1\bar{1}1)$ plane is added (superlattice extrinsic stacking fault, SESF; this is no shear fault). The exact nature of the fault can be determined in the following way. The contrast of the edge fringe (the fringe at the intersection with the foil surface) in bright field images with $s = 0$ depends on the sign of $\sin\alpha$ [1]. If $\sin\alpha > 0$ the fringe is bright and if $\sin\alpha < 0$ the fringe is dark. The values of $\sin\alpha$ are listed in Table 1 for reflections which have been used to image the faults in bright field.

Table 1. Values of $\sin\alpha$ for different reflections and fault vectors

R_f	$\boldsymbol{g} = [002]$	$\boldsymbol{g} = [00\bar{2}]$
$\frac{1}{3}[1\bar{1}1]$	$-\frac{1}{2}(3^{1/2})$	$-\frac{1}{2}(3^{1/2})$
$-\frac{1}{3}[1\bar{1}1]$	$\frac{1}{2}(3^{1/2})$	$-\frac{1}{2}(3^{1/2})$

The experimental observations were that the edge fringe was dark for $\boldsymbol{g} = [002]$ and bright for $\boldsymbol{g} = [00\bar{2}]$, and therefore the fault vector $\boldsymbol{R}_f$ is $+\frac{1}{3}[1\bar{1}1]$. As the foil normal (pointing upwards) is close to [101], $\boldsymbol{R}_f$ points to the upper surface of the foil. The $\boldsymbol{R}_f$ is defined as the displacement of the material below the fault with respect to the material above the fault, and thus the fault corresponds to a removal of one $(1\bar{1}1)$ plane, and the fault is of intrinsic nature, that is, an SISF.

From the direction of the slip traces left behind by the gliding dislocations it was deduced that the $(1\bar{1}1)$ plane of the faults is the same plane as the slip plane of the gliding dislocations. The faults are partially bounded by the intersection of their $(1\bar{1}1)$ plane and the upper surface of the foil and partially by dislocations with Burgers vector $\pm\frac{1}{3}[\bar{1}12]$ (Fig. 31).

The Burgers vector of the dislocations bounding the SISFs has been determined using the $\boldsymbol{g} \cdot \boldsymbol{b}$ invisibility criterion. The invisibility criterion for partial dislocations, such as a dislocation bounding a fault, is slightly different from that for perfect dislocations. For a Burgers vector of the type $\frac{1}{3}\langle 112\rangle$, the criterion for invisibility is $\boldsymbol{g} \cdot \boldsymbol{b} = 0, \pm\frac{1}{3}[1]$ and furthermore (for $s > 0$) $\boldsymbol{g} \cdot \boldsymbol{b} = -\frac{2}{3}$. Care was taken to use only reflections for which the fault is invisible, to prevent any uncertainty regarding visibility of the dislocation. The dislocations were invisible for $\boldsymbol{g} = [220]$ and for $\boldsymbol{g} = [13\bar{1}]$; they were

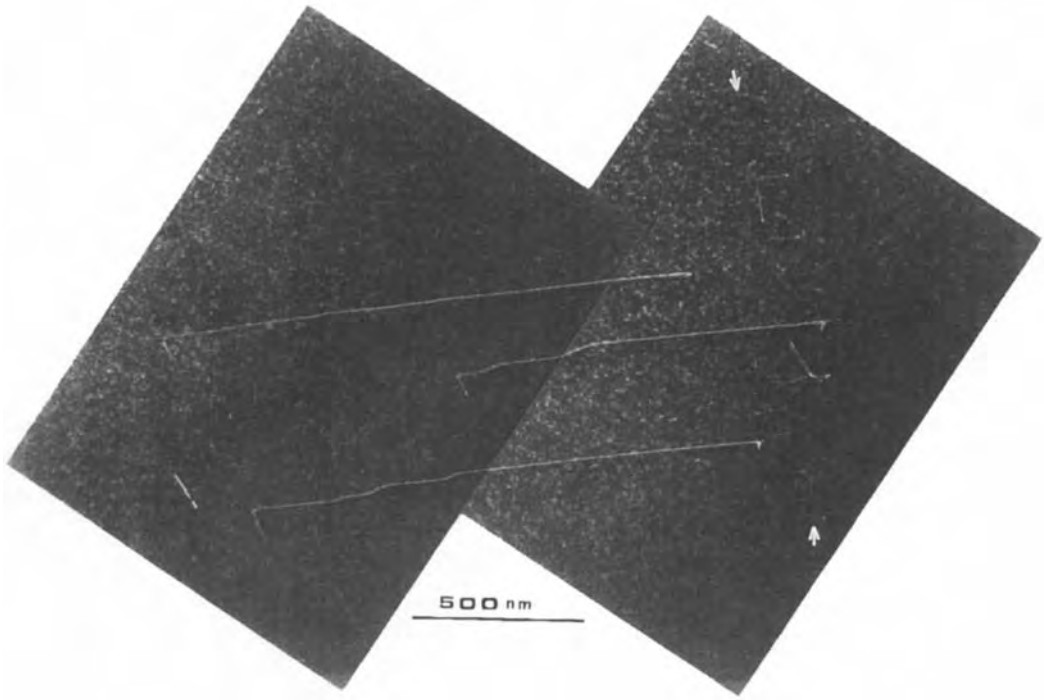


Figure 31. Weak-beam micrograph of the dislocations partially bounding the faults of Fig. 30. Where not bounded by a dislocation, the faults are bounded by the surface of the thin foil. The fringe contrast of the faults is invisible for this imaging condition. The arrows indicate the location of the low-angle grain boundary (cell wall).

visible for $\boldsymbol{g} = [022]$, $\boldsymbol{g} = [1\bar{3}\bar{1}]$, $\boldsymbol{g} = [\bar{1}\bar{1}3]$. The $\boldsymbol{g} \cdot \boldsymbol{b}$ product for dislocations that can bound an SISF is given in Table 2. From Table 2, it can be concluded that the Burgers vector of partial dislocations bounding the SISFs is $\pm\frac{1}{3}[\bar{1}12]$.

The results of the TEM work on the interaction between lattice dislocations with a $\Sigma = 3$ coherent twin boundary show that cracks may emit dislocations parallel to their own plane during propagation. In the case of the sample containing a twin boundary that was analysed in detail, the plane of the crack crossed a coherent twin boundary ahead of the crack, and it might be envisaged that a number of dislocations were emitted from the crack tip in the $(1\bar{1}1)$ plane and

Table 2. $\boldsymbol{g} \cdot \boldsymbol{b}$ product for partial dislocations bounding an SISF on $(1\,\bar{1}\,1)$

$\boldsymbol{b}$	$\boldsymbol{g} = [220]$	$\boldsymbol{g} = [13\bar{1}]$	$\boldsymbol{g} = [022]$	$\boldsymbol{g} = [1\bar{3}\bar{1}]$	$\boldsymbol{g} = [\bar{1}\bar{3}3]$
$\pm\frac{1}{3}[121]$	2	2	2	−2	0
$\pm\frac{1}{3}[21\bar{1}]$	2	2	0	0	−2
$\pm\frac{1}{3}[\bar{1}12]$	0	0	2	−2	2
$\pm\frac{1}{3}[1\bar{1}1]$	0	−1	0	1	1

impinged on the boundary. The slip traces emanating on the other side of the boundary indicate that a number of dislocations have emerged from the boundary in the $(1\bar{1}\bar{1})_{II}$ plane exactly at the point where the plane of the crack intersects with the boundary plane. Thus, it is probable that these dislocations have been emitted from the crack tip and transmitted through the grain boundary. As the [110] direction is common to both grains, the Burgers vector could remain the same in both grains, and no residue is left in the boundary. The line vector is parallel to the intersection of crack plane and the outgoing slip plane, and, thus, transmission could occur without rotation of the dislocation line in the boundary plane. The large number of dislocations may indicate that there was a large force on the leading dislocation of a pile-up in front of the boundary, necessary to cause transmission of the dislocations to the other grain. Grain boundary sources of course cannot be ruled out completely as the origin of the observed dislocations. However, very often, if operation of grain boundary sources is observed, there is generation of dislocations on many different slip planes [105] (our own observations in two other samples indicate the same), while here all the dislocations are on one slip plane.

As low-angle grain boundaries, also called cell walls, consist of a dislocation network, the interaction between gliding lattice dislocations and low-angle grain boundaries is essentially a dislocation–dislocation interaction. In order to find an explanation for the formation of the SISFs, it is useful to know the type of the dislocations that were gliding in the slip band. Therefore, a number of dislocations showing such a configuration, so that it could be expected that they had been gliding in the slip band, were analyzed. Most of the dislocations found in the slip were screw dislocations with a Burgers vector of [011]. We may assume that the gliding dislocations in the slip band had a Burgers vector of [011]. This assumption can be supported as follows. If we suppose that the $\frac{1}{3}[\bar{1}12]$ dislocations bounding the SISFs were created by the gliding dislocations, the Burgers vector of the gliding dislocations could either be [011] or $[\bar{1}01]$, as only these dislocations are glissile in the $(1\bar{1}1)$ plane and at the same time can have $\frac{1}{3}[\bar{1}12]$ as one of their dissociation products. If we consider the tensile axis in the thin region to be the same as the macroscopic tensile axis, the Schmid factor for the [011] dislocations in the $(1\bar{1}1)$ plane is 0.38, while the Schmid factor for the $[\bar{1}01]$ dislocations is 0.15. Therefore, we can conclude that the gliding dislocations had a Burgers vector of [011]. As the dislocations in the slip band that were analyzed had mostly screw character, we may assume that the gliding dislocations also had screw character. Also, it can be expected that the gliding dislocations were dissociated into two $\frac{1}{2}[011]$ superpartials, which is the dissociation that is usually found in Ni_3Al. A schematic overview of the configuration is given in Fig. 32.

Now, we consider the formation mechanism of the SISFs in more detail. A gliding dislocation intersects a dislocation in the cell wall. Although the dissociation of the perfect dislocation in the cell wall into superpartial dislocations is not clearly visible, we may assume that it is dissociated in the cell wall into two $\frac{1}{2}[1\bar{1}0]$ partials. Because of the intersection, two jogs (steps in the dislocation line; the steps do not lie in the slip plane) of $\frac{1}{2}[1\bar{1}0]$ have

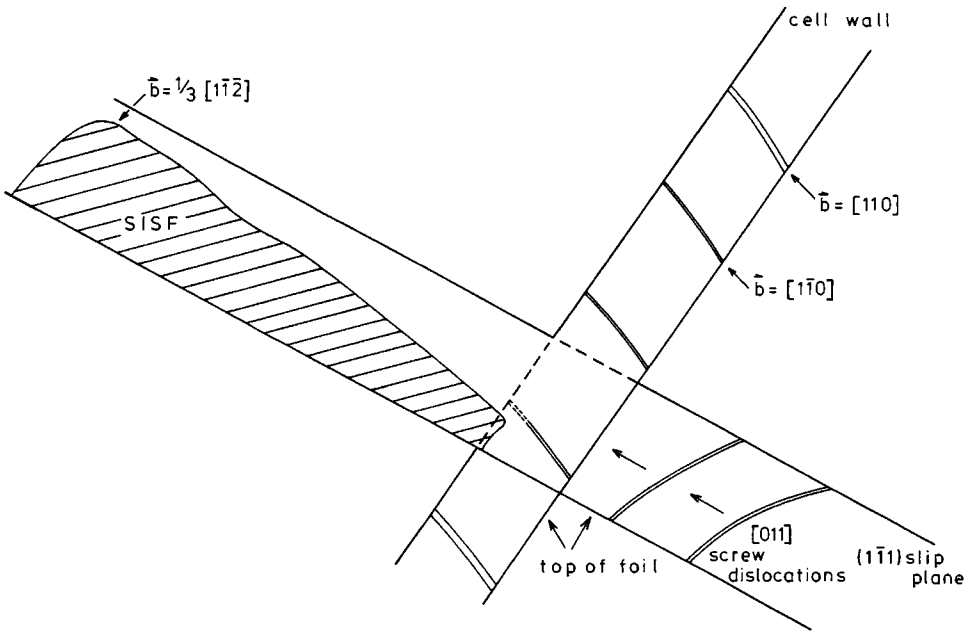


Figure 32. Schematic view of the gliding dislocations, the low-angle grain boundary (cell wall) and the SISF that is formed. The intersections of the slip plane of the gliding dislocations with the upper and lower surfaces of the thin foil are indicated. Also, the intersection of the grain boundary plane and the foil surfaces has been indicated.

been created in the line of the gliding dislocation. As the separation between the two $\frac{1}{2}[1\bar{1}0]$ partials in the cell wall is very small, the jogs are very close to each other. The jogs cannot glide in the $[2\bar{1}1]$ direction of motion of the [011] screw dislocation in the $(1\bar{1}1)$ plane, and therefore the motion of the jogs can only occur by climb, and as this is a slow process the motion of the dislocation is hindered at the jogs, that is, the dislocation is pinned by the jogs. The dislocation will bow out at both sides of the pinning point under the influence of the stress. The leading $\frac{1}{2}[011]$ superpartial dislocation can be allowed to move under the applied stress in the following way: if a $\frac{1}{6}[21\bar{1}]$ Shockley partial dislocation moves from the trailing to the leading superpartial, the antiphase boundary (APB) separating the $\frac{1}{2}[011]$ partials is transformed into an SISF, which has much lower energy. By the exchange, the trailing $\frac{1}{2}[011]$ superpartial transforms to a $\frac{1}{3}[\bar{1}12]$ dislocation partially bounding the SISF (Fig. 32).

In this way, the force keeping the two $\frac{1}{2}[011]$ dislocations together because of the high APB energy is greatly reduced as the SISF energy is much lower than the APB energy and, thus, the leading $\frac{1}{2}[011]$, together with the $\frac{1}{6}[21\bar{1}]$ can proceed, and then a large region if SISF can be formed. It might be envisaged that the termination of the SISF region develops in a similar way as the loop-pinching mechanism described by Pak et al. [106]. In our case, we are dealing with a partial loop that is bounded by the foil surface and no next

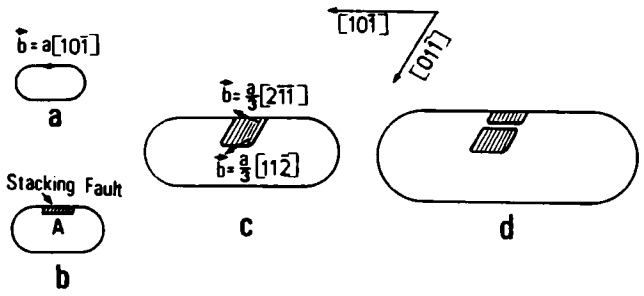


Figure 33. The SISF formation mechanism proposed by Pak et al. [106]. Part of an expanding perfect dislocation loop is dissociated into two $\frac{1}{3}\langle 112\rangle$-type partials with an SISF in between. As the perfect loop expands further, the separation between the two partials increases, until the SISF is pinched off.

loop of SISF is created, but instead a new [011] perfect dislocation might be formed, which proceeds in the $(1\bar{1}1)$ plane. Another way, differing in detail, of describing the generated SISF is by a dissociation of the leading superpartial into $\frac{1}{6}[21\bar{1}]$ and $\frac{1}{3}[\bar{1}12]$. The superpartial $\frac{1}{3}[\bar{1}12]$ is repelled from the Shockley partial, which in turn may react with the trailing $\frac{1}{2}[011]$ dislocation to form $\frac{1}{3}[121]$ [107, 108].

There have been a number of other observations of SISFs in ordered alloys such as Ni_3Al. In Ni_3Al, Baker and Schulson [109] observed pairs of $\frac{1}{3}\langle 112\rangle$ dislocations with an SISF inbetween, separated by around 100 nm. The $\frac{1}{3}\langle 112\rangle$ pairs had parallel Burgers vectors and antiparallel line vectors. These pairs were thought to be elongated dislocation loops that were truncated because of the thinning process during preparation of the foil. Some observations were made of rows of loops of SISF that were several hundreds of nanometers long. Pak et al. [106] made observations in Ni_3Ga, which is a very similar material to Ni_3Al, of widely extended SISFs with lengths of several hundreds of nanometers to a few micrometers, elongated along $\langle 110\rangle$ directions and bounded by $\frac{1}{3}\langle 112\rangle$-type partial dislocations. Veyssiere et al. [108] made observations of two $\frac{1}{2}\langle 110\rangle$ superpartials that were dissociated in a $\{100\}$ plane, connected by a ribbon of APB. Parts of one of the $\frac{1}{2}\langle 110\rangle$ dislocations had dissociated in an inclined $\{111\}$ plane into a $\frac{1}{6}\langle 112\rangle$-type edge Shockley partial and a $\frac{1}{3}\langle 112\rangle$ type partial dislocation. These two dislocations were connected by a ribbon of SISF on the $\{111\}$ plane.

Several explanations have been put forward for the SISF formation. Pak et al. [106] explained the formation of loops bounding SISFs by means of an expanding perfect dislocation loop (Fig. 33), which partially dissociates into two $\frac{1}{3}\langle 112\rangle$ partial dislocations. Upon further expansion of the perfect loop, the SISF is elongated and, finally, part of the SISF is pinched off, and a faulted loop containing the SISF is created. Kear et al. [110] discussed mechanisms by which gliding dislocations on intersecting slip planes could react, and they showed that an SISF is formed in certain reactions. The mechanism described by Kear et al. is essentially a dislocation interaction mechanism involving two different slip planes, while the mechanism described by Pak et al. is essentially a single slip mechanism. Pak et al. do not go further into the reason for the initial dissociation of part of the perfect dislocation loop into $\frac{1}{3}\langle 112\rangle$ partial dislocations. The SISF formation in our observations resembles the mechanism of Pak et al., as

only one slip plane is involved. The creation of jogs in the line of the gliding dislocations because of the intersection with the dislocations in the cell wall could be the reason for the dissociation into $\frac{1}{3}\langle 112\rangle$ partials under the influence of the applied stress.

Observations of the same dislocation–grain boundary system were made in type 304 stainless steel by Clark and Wagoner [111] and in 310 stainless steel by Lee et al. [112]. In the 304-type stainless steel, a configuration in lightly deformed material was observed which showed a pile-up of dislocations on one side of the boundary and dislocations which appeared to have emerged on the other side. It was concluded that transmission through the twin boundary had taken place. Lee et al. observed the same mechanism of transmission for the same type of dislocations through the same boundary in their in situ deformation experiments in 310-type stainless steel.

There are some special boundaries, such as a coherent $\Sigma 3$ boundary as investigated here, which at first sight would not act as strong barriers. This certainly is so in the case of $\Sigma 3$ for polycrystalline f.c.c. metals where the slip planes in both grains are symmetrically disposed and have a common line of intersection with the grain boundary plane. Lattice dislocations of the type $\frac{1}{2}\langle 110\rangle$ may transmit directly from one grain to the other or by the generation of a residual $\frac{1}{6}\langle 112\rangle$ dislocation in the grain boundary plane itself. It is quite clear from the analysis mentioned above that the situation is far more complex in $L1_2$ ordered alloys, for example. This has to do with both the fact that we have to deal with superlattice dislocations, that is, two unitary $\frac{1}{2}\langle 110\rangle$ lattice dislocations bounding an APB as well as the $\Sigma 3$ boundary structure in ordered compounds after the interaction with the complete Burgers vector being quite different from the situation in pure f.c.c. metals. Tichelaar [113] made a TEM analysis of the various grain boundary structures that arise after the interaction of superlattice dislocations with a coherent twin boundary in $L1_2$. When an APB terminates in a coherent twin boundary it causes a translation of $\frac{1}{2}\langle 110\rangle$ of the two crystals with respect to each other, that is, the boundary may change from a symmetrical configuration (S) into an asymmetrical one (A). These changes between S and A may also occur by a glissile superlattice dislocation or by a grain boundary dislocation of the type $\frac{1}{6}\langle 112\rangle$. It should be realized that there are three crystallographic variants of type A, meaning that either Shockley partial dislocations in the boundary or a terminating APB may also cause a transformation among the A structures themselves. Of course, the energies of the various A structures should be identical based on symmetry arguments. In Tables 3 and 4 some energies of these faults are listed, calculated using N-body potentials [114].

Table 3. Calculated planar fault energies ($mJ\,m^{-2}$) for different shear faults [114]

Planar fault	APB on (001)	ABP on $(1\bar{1}1)$	CSF on $(1\bar{1}1)$	SISF on $(1\bar{1}1)$
Translation vector	$\frac{1}{2}[110]$	$\frac{1}{2}[110]$	$\frac{1}{6}[121]$	$\frac{1}{3}[21\bar{1}]$
Cu_3Au	55	56	40	16
Ni_3Al	53	227	191	12

Table 4. Calculated grain boundary energies ($mJ\,m^{-2}$) for different ordering states after relaxation. The cubic sublattices containing the B atoms are indicated by their relative translation vector (in the coordinate system of the upper grain). The relaxed configuration with full symmetry (i.e., B atoms are mirror related to B atoms) is indicated by S [114]

Ordering state	Materials	S	A1	A2	A3
Translation vector		[000]	$\frac{1}{2}$[110]	$\frac{1}{2}$[101]	$\frac{1}{2}$[011]
$\Sigma 3$	Cu_3Au	7	47	47	47
	Ni_3Al	6	208	208	208
$\Sigma 11$	Cu_3Au	363	344	358	359
	Ni_3Al	506	490	605	606
$\Sigma 27$	Cu_3Au	697	701	737	737
	Ni_3Al	996	994	1105	1105

It turns out that eight different configurations are possible when a superlattice dislocation with total Burgers vector ⟨110⟩ is absorbed in a {111} coherent twin boundary (Fig. 34). In fact, the self energy of the grain boundary dislocations can be lowered by the dissociation into Shockley partials which can be easily anticipated by looking at the various energies involved as compiled in Table 4. It should be realized that here we have assumed that the superlattice dislocation enters the boundary structure at the S configuration. However, if it arrives at the grain boundary in an A configuration, different configurations may arise depending which of the A structures is actually involved; for example, A1 is changed into A3 but the energy is unaltered, meaning that the absorption of lattice dislocation is favorable since the APB in the bulk is removed. However, A2 may also be transformed upon interaction into a symmetrical S structure, that is, lowering its energy (Table 4), but is transformed back again upon absorption of the trailing unitary $\frac{1}{2}$⟨110⟩ dislocation.

It is interesting to compare the in situ observations of Ni_3Al with those of Cu_3Au [112]. The micrographs displayed

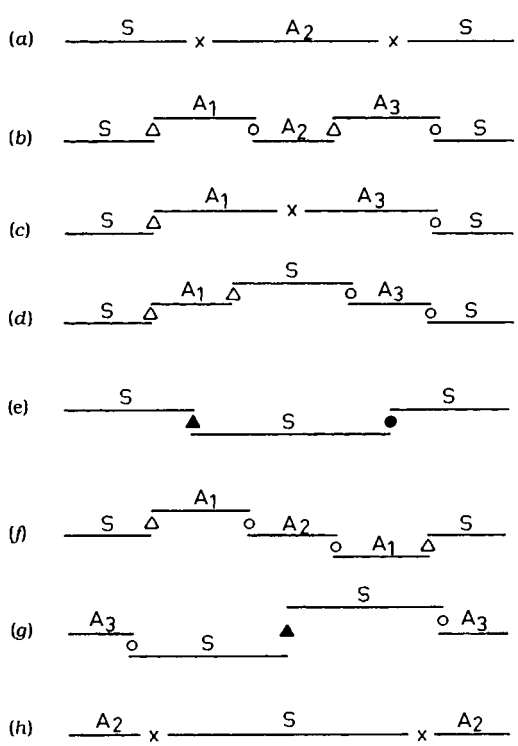


Figure 34. Different dissociation schemes of a [101] superlattice dislocation in $L1_2$ structure absorbed in a $(11\bar{1})$ coherent twin boundary. Superpartials $\frac{1}{2}$[101] (×) may dissociate into Shockley partial dislocations with Burgers vectors $\frac{1}{6}[2\bar{1}1]$ (△) and $\frac{1}{6}$[112]. ○ and ▲ refer to DSC dislocations with $\frac{1}{3}$[112] and $\frac{1}{3}[2\bar{1}1]$, respectively.

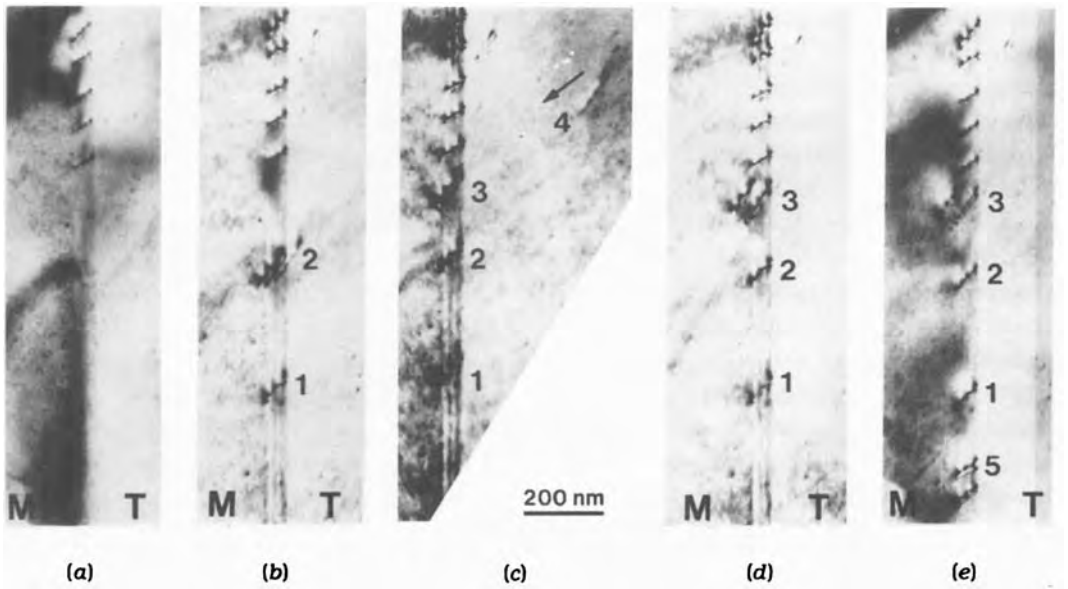


Figure 35. Dark field TEM images. Upon increasing strain, additional superlattice dislocations arrive at the twin boundary.

in Fig. 35 show the dislocation activity after incremental deformation. Following a large strain (Fig. 35e, position 5), a dislocation pair was observed to move. This dislocation pair was present in the A structure whereas the other dislocation pairs were present in the S structure. It can be concluded that most of the superlattice dislocations moving in T were absorbed into the boundary plane. Also, transmission was observed into M, leaving slip traces in a direction parallel to the slip plane. On increasing deformation, dislocations were seen to move slowly into T. In this case the dislocation traveled as the usual type, which can be confirmed from the APB contrast (Fig. 36). In Fig. 37 another view of the boundary structure is presented. Here, the boundary is imaged using weak-beam common fundamental reflection, showing dislocation pairs, as well as using a common superlattice reflection showing APB-type contrast between the dislocation pairs. It can be concluded that before the superlattice dislocation arrived, the boundary structure was S but has altered locally to A.

In these experiments, it was deduced from examinations of the activated slip systems that the symmetric slip plane in the other grain had the highest Schmid factor. Although there were a number of observations of transmission, absorption in the boundary plane was also observed. A possible explanation of the tendency for absorption could be found in a lower value of the energy of the ordering fault in the grain boundary plane between the two $\frac{1}{2}[110]$ superpartial dislocations compared with the value of the APB between the superpartials in the bulk. Computer simulations show that for Ni_3Al this energy difference is $25\,mJ\,m^{-2}$ (for Cu_3Au it is $16\,mJ\,m^{-2}$) [114]. Thus, regarding this energy difference, absorption in the boundary might be expected.

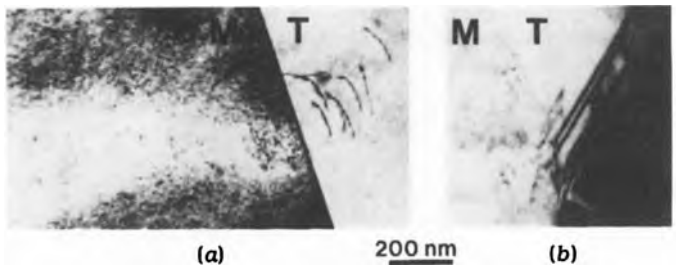


Figure 36. Bright field TEM images of the same area as depicted in Fig. 35 using (a) a fundamental reflection and (b) a superlattice reflection.

It turned out that the mechanism of the interaction between lattice dislocations and grain boundaries is similar for boundaries in ordered and disordered (or f.c.c.) materials [114]. In the case of absorption into grain boundary dislocations in ordered material, during movement of the grain boundary dislocation an ordering fault is left behind in the boundary plane, and anti-site defects may be created. Because of these phenomena, movement of grain boundary dislocations in ordered material is hindered, and, in the case of extended slip, stress concentrations will develop at the boundary upon arrival of more lattice dislocations at the same location in the boundary plane. In the case of transmission, computer simulations [114] show that transmission through an ordered boundary occurs at high stress

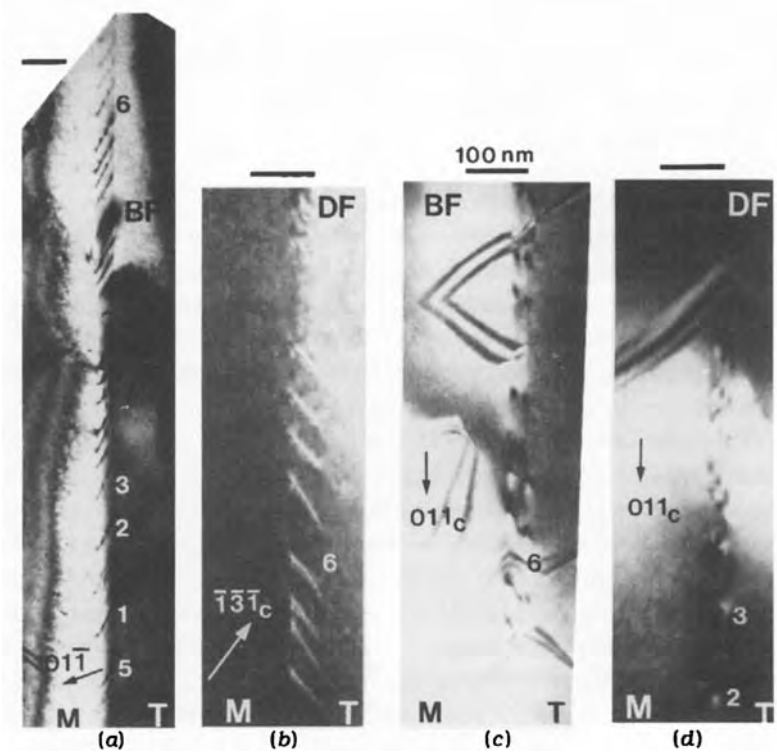


Figure 37. Bright field and dark field TEM images of the coherent twin boundary. (b) Dislocation pairs are imaged using a weak-beam fundamental reflection. (c, d) A-type structure is imaged as APB contrast. (d) The area shown here is directly below the areas shown in (b) and (c).

levels and the two superpartial dislocations constituting the arriving dislocation will decrease their separation in response to the applied stress, when the leading superpartial dislocation is halted at the boundary. In this way, stress concentrations near the boundary will be generated. The effects mentioned above are greater for higher ordering tendency. The results obtained indicate that the interaction between lattice dislocations and grain boundaries, as observed by TEM and computer modeling, is essential for the explanation of the intergranular fracture occurring in a number of ordered compounds, in particular those with a high ordering energy.

In this view, the mechanism by which the ductility of hyperstoichiometric (i.e., nickel-rich) Ni_3Al is increased by the addition of small amounts of boron could be explained as follows. It has been established that boron segregates to the grain boundaries in Ni_3Al [115, 116]. Nickel atoms are thought to cosegregate with the boron atoms [117–119], thus causing a nickel-rich grain boundary. In this way, the long-range order could be disturbed in a region very close to the grain boundary. Although there is still some debate, there is experimental evidence that there indeed exists a disordered region at grain boundaries in nickel-rich Ni_3Al [120, 121]. As shown in the simulations [114] in f.c.c. material, which is essentially 'disordered $L1_2$', the grain boundaries in disordered material are much more transparent for dislocations. As a consequence of the 'disordered region' in Ni_3Al by boron additions, the net result may be that slip can be transmitted via the nucleation of superlattice partial dislocations under a severely reduced stress. The reduction will be of the order of γ_{APB}/b (size of the Burgers vector), that is, about 900 MPa (ignoring the lattice hardening due to boron strengthening effects [122]).

Finally, one should be careful in generalizing these results obtained for all $L1_2$ ordered systems. Indeed, as the mechanical properties of more $L1_2$ materials have been investigated experimentally, it has been found that several of them are intrinsically brittle because of intergranular failure, similarly to Ni_3Al. However, it has to be emphasized that in intermetallics, as in Al_3Ti, cleavage fracture is also found to be the failure mode, rather than intergranular failure. In cleavage fracture it is feasible that a mechanism for crack tip plasticity exists that does not require macroscopic dislocation motion at all [123]. Local shears produced at the crack tip which reverse once the crack has passed by is one of the possible mechanisms.

Particle-Hardened Materials

Direct observations of crack propagation and crack tip deformation are also possible via in situ TEM experiments. To be somewhat consistent with the previous experiments we describe as an example some in situ experimental results on the deformation characteristics of superalloys, that is, nickel-base matrix containing Ni_3Al–$L1_2$ ordered precipitates (MA 6000 and PM Astroloy). The common feature of nearly all nickel-base superalloys is a dispersion of an f.c.c. intermetallic phase, called γ' in an f.c.c. solid solution γ. Although γ' is based on Ni_3Al, it may contain substantial amounts of other alloying elements such as titanium, cobalt, tantalum and molybdenum [124]. The coherently ordered γ' is the most important hardening phase in

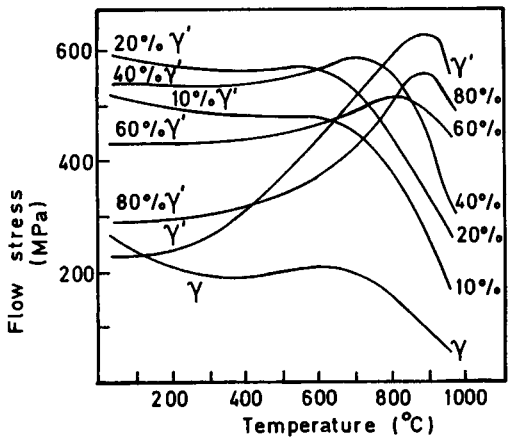


Figure 38. The temperature dependence of the 0.2% flow stress for Ni–Cr–Al containing different volume fractions of γ'.

superalloys. In the cubic morphology, γ' precipitates are aligned along preferred directions such that the faces of the cubes are parallel to $\{001\}\gamma$. Figure 38 shows the exceptional temperature dependence of the flow stress of γ', which increases strongly from room temperature to 900°C, after which the stress rapidly falls off [125]. A successful theory to account for this effect is that of Takeuchi and Kuramoto [126], in which dislocations on {111} planes are assumed to become locally pinned by thermally activated cross-slip onto {001} planes, a process driven by the anisotropy of the APB energy and aided by the resolved shear stress for cubic slip. The qualitative explanation is that an increase in temperature promotes cross-slip, and therefore a higher flow stress. The decrease in flow stress at higher temperature is contributed to macroscopic slip on {001} planes. Alloys with large fractions of γ' reflect the pure γ' behavior, and about 50% γ' gives the best overall properties. For some information on the mechanical properties of these two phase systems, see

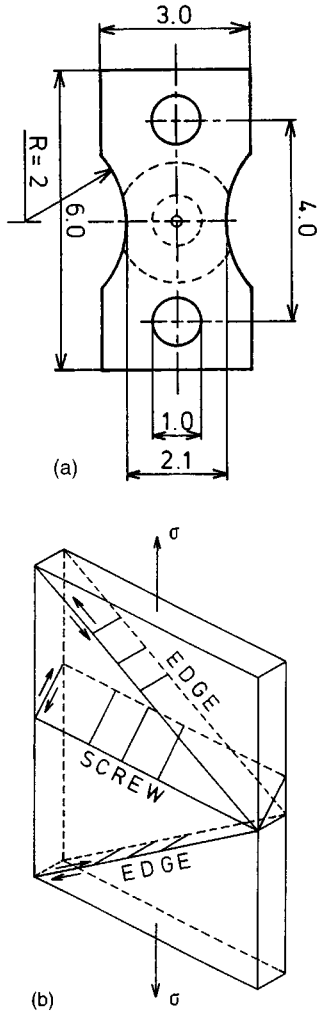


Figure 39. (a) Specimen for in situ observations and (b) the slip geometry expected.

Heredia [127]. Let us return to the possibilities of in situ crack propagation and crack tip deformation in TEM.

A typical specimen for the in situ observations is shown in Fig. 39. The specimen was cut by spark erosion and ground to a thickness of 130 μm. The final configuration was reached by electropolishing in two steps: first thinning of the

gauge section with a diaphragm of 2.5 mm and subsequently polishing with a diaphragm of 1 mm until perforation. Ohr [128] has analyzed the configuration in Fig. 39, and has calculated the planes of maximum shear stress, which are depicted in Fig. 39b. To minimize their line length, dislocations lie across the thickness of the sheet. The dislocations on planes 1 and 3 are predominantly of the edge type, while those on plane 2 are predominantly of screw type. When edge dislocations are emitted simultaneously on the two inclined slip planes, the resulting Burgers vectors of these dislocations are in the direction parallel to the tensile axis, and the induced crack opening displacement is parallel to the tensile stress (mode I). Mode II crack tip deformation occurs if edge dislocations are emitted on one of the inclined planes. Screw dislocations on slip plane 2 give rise to mode III crack tip deformation. The three modes of fracture and crack tip deformation are depicted in Fig. 40. The choice of the slip system is dictated by the shear stress close to the crack tip, whereas the bulk Schmid factor influences dislocation behavior in the region outside the crack tip area. In fact, Ohr has found no case of a pure single mode of loading, and among the three modes of deformation,

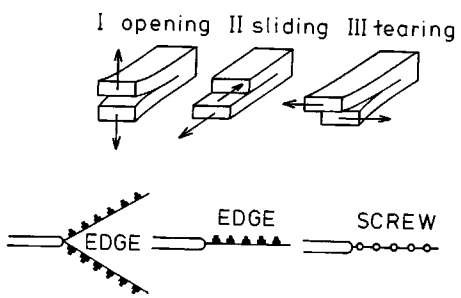


Figure 40. The three modes of fracture.

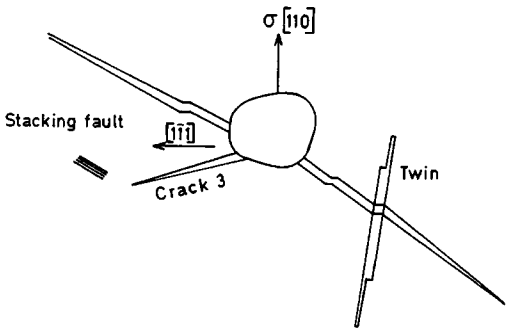


Figure 41. Schematic representation of the crack configuration.

mode III was favored over mode II and in turn mode II was preferred over mode I.

At room temperature, mechanical stress was applied to a sample of MA 6000 with longitudinal grains in the direction of the applied stress. Two cracks were initiated at the edge of the hole and propagated as indicated in Fig. 41, at about 55° to the tensile axis, which is the longitudinal [110] grain direction. By cyclic deformation at about 600°C a third crack was initiated. The two directions of crack propagation coincide with the orientation of twins; a third twin orientation has been observed as well.

In thin areas near the hole a crack propagates in a straight line. The propagation of a crack across a twin in a thin area has been scrutinized. Before a crack propagates across a twin a plastic zone is built up on the opposite side of the twin. Finally, the crack cuts through the twin in a different direction and propagates on the other side parallel to the original direction, resulting in a step in the crack surface. In Fig. 42 the part of the twin below the crack is out of contrast. In thicker areas, crack propagation occurs by forming spherical cavities in front of the crack tip. The cavity increases in size, and at a further stage the

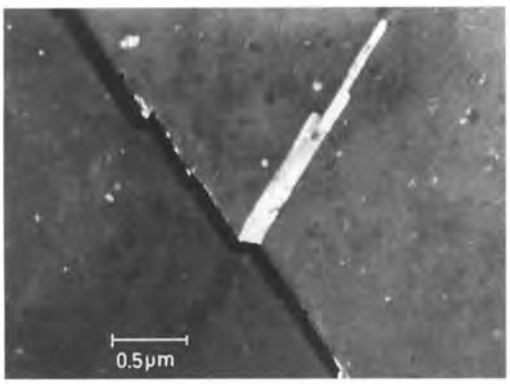


Figure 42. A crack through a twin.

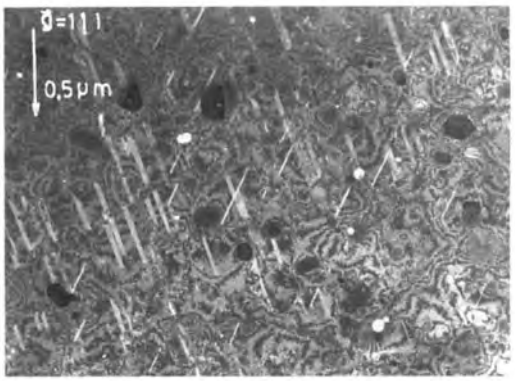


Figure 44. Stacking faults inside γ' precipitates.

crack grows until it reaches the cavity. In front of the third crack tip, stacking faults are located at a distance of about 3 μm (see Fig. 43). The orientation difference between the crack and faults is 46°. These stacking faults lie in the direction of the primary cracks. Due to the slip of partial dislocations the fault contrast disappears, but returns as expected in the case of the growth of a microtwin. The growth of a stacking fault inside a precipitate has been observed as well. Besides partial dislocations, as depicted in Fig. 43, slip traces were also observed in the direction of the crack at a distance of 5 μm. It seems that the third crack propagates by a combination of the opening mode with either the sliding mode or the tearing mode. As far as the primary cracks are concerned no evidence was found for the opening mode.

The in situ straining experiments show that at low and intermediate temperatures, that is, below about 800°C, MA 6000 deforms by the glide of dislocations in planar arrays. As observed in situ, partial dislocations cause stacking faults (Fig. 44) in the precipitates, and successive dislocations produce microtwins (Fig. 45). These twins have also been found after fracture

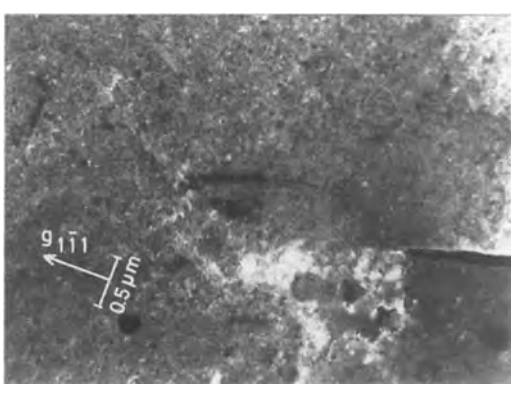


Figure 43. Stacking faults in front of a crack tip.

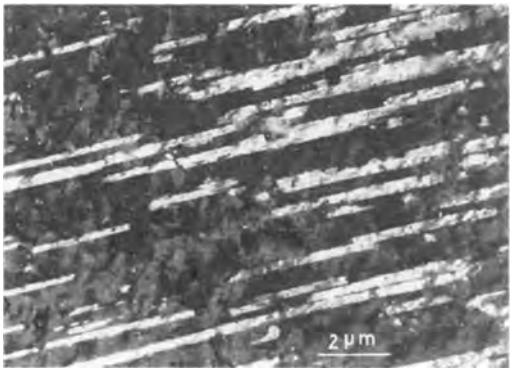


Figure 45. Twins after deformation until fracture at 760°C.

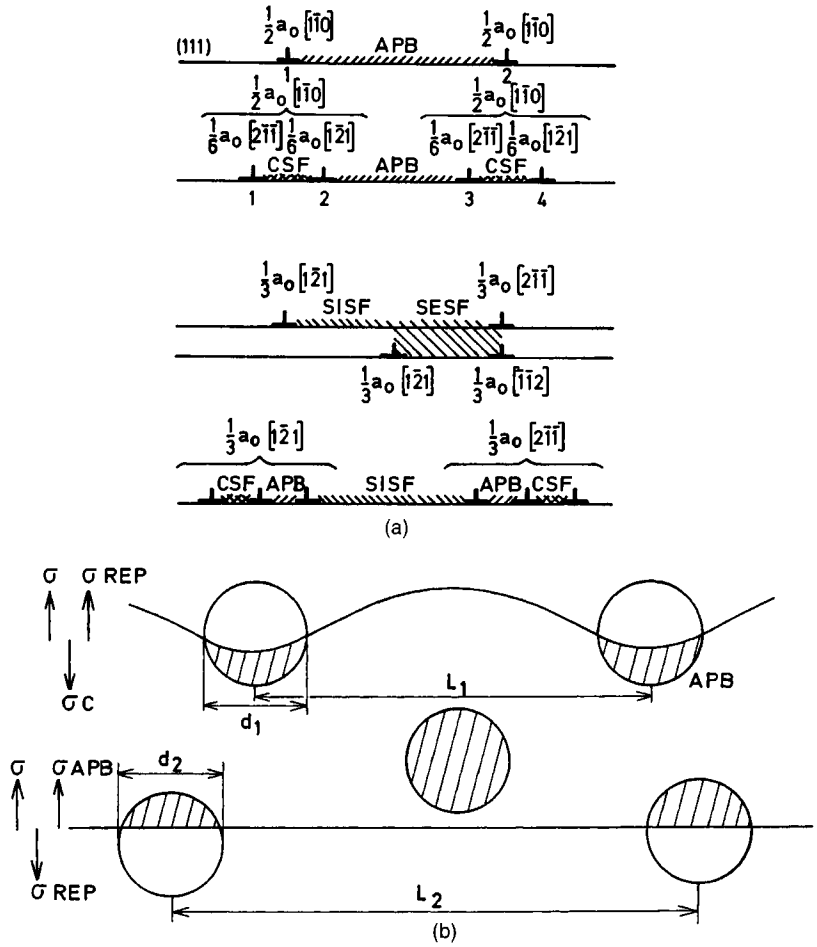


Figure 46. (a) Dissociation of superlattice dislocations in $L1_2$. (b) Dislocation pair in its glide plane.

at room temperature but in a lower density than after fracture at around 800°C. In a typical twin [011] diffraction pattern a $[1\bar{1}1]$ twin axis has been found, proving that the twin plane is an octahedral plane.

The main dissociation mechanisms of unitary dislocations in $L1_2$ ordered materials were described at the beginning of this section (see also Table 3 and Fig. 46A). Here the mechanisms will be applied to $L1_2$ ordered precipitates. In the case of a low APB energy, dislocation motion is not seriously hindered by precipitates. To shear the precipitates no rearrangement of the dislocation is necessary. If the matrix stacking fault energy is high, unit dislocations are not dissociated in the matrix, and shearing of the precipitates and matrix will occur by single-unit dislocations or by paired dislocations according to the physical picture depicted in Fig. 46B, where the shear stress is given by [129]

$$\sigma = \frac{\gamma_{APB}}{2b}\left(\frac{d_1}{L_1} - \frac{d_2}{L_2}\right) \tag{18}$$

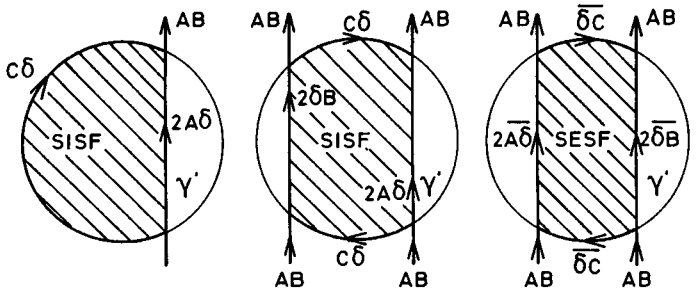


Figure 47. Top view of SISFs and an SESF in γ' precipitates.

If the matrix stacking fault is low, precipitates and matrix will be deformed by partial dislocations which create stacking faults in the matrix and complex stacking faults (CSFs) in the precipitates. In the present case of the nickel-base superalloys, both the stacking fault energy and APB are considered to be high. So, the matrix will be deformed easily by unit dislocations that are halted at the interfaces between the matrix and precipitates. An ordering fault can be avoided if the precipitates are sheared by $2A\delta$ (Thompson notation [130, 131]), depicted in Fig. 47. A second dislocation confines the stacking fault (Fig. 47b). The same mechanism is applicable for extrinsic faults (Fig. 47c) [132, 133]. The dissociation of AB is also possible into Shockley partial dislocations according to $2AB \rightarrow \bar{A}\bar{\delta} + A\delta + \delta B + \delta\bar{E}$ (Fig. 48). In this case $2\delta\bar{B}$ is coupled to $2\bar{A}\bar{\delta}$ by a CSF–APB fault that produces a highly constricted configuration. However, a relaxed mode of dissociation is reached if $\delta\bar{B}$ is separated from δB, which is also located between $\bar{A}\bar{\delta}$ and $A\delta$ [134]. Then the double fault in Fig. 48a is transformed into an SISF (Fig. 48b). The dipole displacement is indicated by the symbol $\diamondsuit$, and the dissociation has the following form:

$$2AB \rightarrow \bar{A}\bar{\delta} + \diamondsuit + A\delta + \delta B + \diamondsuit + \delta\bar{E}.$$

Deformation of precipitates by partial dislocations according to Fig. 48b is energetically favorable in comparison with Fig. 48a because of the lower energy of an SISF relative to the CSF–APB fault. The reduced interaction energy must offset the higher core and elastic energies of the

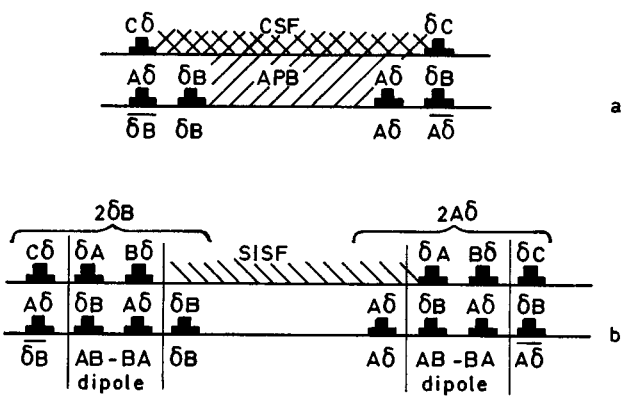


Figure 48. Dissociation of 2AB without and with dipole rearrangements.

dislocations. The difficult step is clearly associated with the formation of a dipole at the interface between matrix and precipitate. This dissociation scheme seems to be appropriate only for high-temperature deformation where thermal vacancies may interact with dislocation cores to provide the required dipole displacements. This slip mode is termed viscous slip because the necessary vacancy migration is thermally activated. The hard sphere model of Strutt et al. [135] clearly shows the effect of such a dipole displacement.

If the matrix stacking fault energy is low enough to make dissociation of unit dislocations feasible, deformation will no longer occur by unit dislocations because the two total superlattice dislocations with Burgers vectors at 60° can react to form intrinsic–extrinsic fault pairs. These fault pairs with a net Burgers vector of $6A\delta$ have been observed in precipitation-hardened alloys as well [136], but will not be considered here since dissociated matrix dislocations were not observed by CTEM in the present case.

An experimental TEM method to verify whether the faults inside γ' precipitates are CSFs or SISFs is to determine the Burgers vector of the leading dislocation of a fault. The analysis is along the same lines as listed in Table 2. Here the determination of the Burgers vector is performed with the invisibility criterion $\boldsymbol{g}\cdot\boldsymbol{b} = 0, \pm\frac{1}{3}, -\frac{2}{3}, \frac{4}{3}$. In all cases the deviation from the Bragg reflection is taken as positive ($s > 0$). The results are summarized in Table 5. The invisibility for $\boldsymbol{g}\cdot\boldsymbol{b} = +\frac{4}{3}$ is ambiguous. Nevertheless, comparison of Fig. 49a ($\boldsymbol{g}\cdot\boldsymbol{b} = +\frac{4}{3}$) with Fig. 49b ($\boldsymbol{g}\cdot\boldsymbol{b} = -\frac{4}{3}$) reveals a clear difference in contrast. Care was taken on reversing $\boldsymbol{g}$ to maintain the

Table 5. $\boldsymbol{g}\cdot\boldsymbol{b}$ and $\boldsymbol{g}\cdot\boldsymbol{R}$ analyses of the fault configurations shown in Figs. 49 and 50 (i, invisible; v, visible). The extinction criteria for a stacking fault are $\alpha = n\pi$.

$\boldsymbol{g}$	$\boldsymbol{g}\cdot\boldsymbol{b}$	Observation	$\alpha/\pi = 2\boldsymbol{g}\cdot\boldsymbol{R}$ $R = \frac{1}{3}[11\bar{2}]$	Observation	Reference
	$b = (a/3)[11\bar{2}]$				
[220]	$\frac{4}{3}$	i	$\frac{8}{3}$	v	Fig. 49a
$[\bar{2}\bar{2}0]$	$-\frac{4}{3}$	v	$-\frac{8}{3}$	v	Fig. 49b
$[\bar{2}00]$	$-\frac{2}{3}$	i	$-\frac{4}{3}$	v	
[200]	$\frac{2}{3}$	v	$\frac{4}{3}$	v	
$[20\bar{2}]$	2	v	4	i	
[020]	$\frac{2}{3}$	v	$\frac{4}{3}$	v	
$[0\bar{2}0]$	$-\frac{2}{3}$	i	$-\frac{4}{3}$	v	
	$b = (a/6)[1\,-2\,-1]$				
$[0\bar{2}\bar{2}]$	1	v			Fig. 50a
$[\bar{2}\bar{2}0]$	$\frac{1}{3}$	i			
$[\bar{2}00]$	$-\frac{1}{3}$	i			
$[\bar{1}\bar{1}\bar{1}]$	$\frac{1}{3}$	i			
$[\bar{2}02]$	$-\frac{2}{3}$	i			
$[\bar{1}11]$	$-\frac{2}{3}$	i			Fig. 50b
$[1\bar{1}\bar{1}]$	$\frac{2}{3}$	v			
$[\bar{1}\bar{1}1]$	0	i			

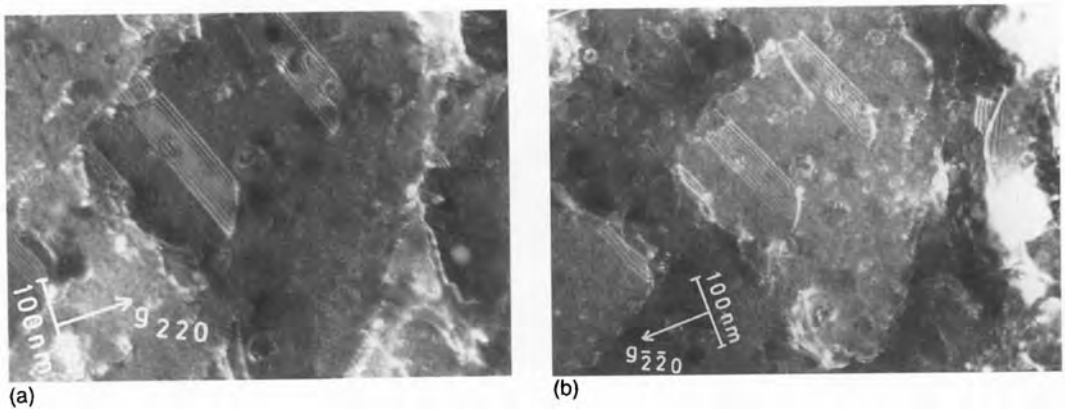


Figure 49. $(a/3)[11\bar{2}]$ dislocation in front of an SISF inside γ'. (a) $\boldsymbol{g}\cdot\boldsymbol{b} = +\frac{4}{3}$ and (b) $\boldsymbol{g}\cdot\boldsymbol{b} = -\frac{4}{3}$ (see Table 5).

same value for the deviation of the Bragg condition. The occurrence of an $a/6[11\bar{2}]$ partial dislocation is excluded because of the visibility of the leading dislocation of the fault for $\boldsymbol{g} = [\bar{2}\bar{2}0]$. From Table 5 it may be concluded that an $(a/3)[11\bar{2}]$ partial dislocation creates the fault inside the γ' precipitate. By observing the outermost fringes, the intrinsic character of the faults has been proven. The dislocation fault configuration shown in Fig. 49 is just one example of a large number with the same extinction behavior. Along the fringes of some superlattice intrinsic stacking fault a dislocation segment can be resolved. In Fig. 50 such a segment is visible, where it can be seen that this fault extends over the whole γ' precipitate. Here again with the invisibility criterion the character of the dislocation can be determined. The Burgers vector is $\frac{1}{6}[1\bar{2}\bar{1}]$ (see Table 5). Although the plane of the fault has been determined by tilting, the exact value of the displacement is unknown. The three possible values of $\boldsymbol{R}$ are: $\frac{1}{3}[\bar{2}1\bar{1}]$, $(a/3)[112]$ and $(a/3)[1\bar{2}\bar{1}]$. The third would imply that $(a/2)[1\bar{2}\bar{1}]$ is operative, which is rather unlikely

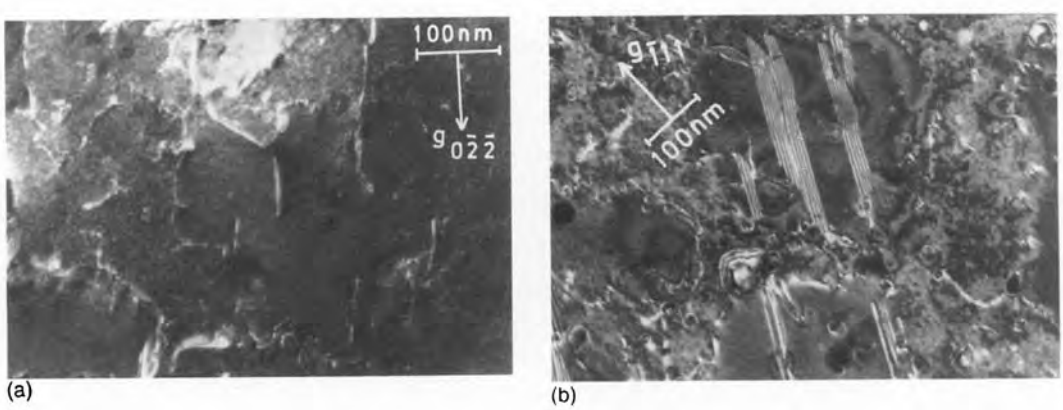


Figure 50. $(a/6)[1\bar{2}\bar{1}]$ dislocation along the fringes of an SISF. (a) $\boldsymbol{g}\cdot\boldsymbol{b} = +1$ and (b) $\boldsymbol{g}\cdot\boldsymbol{b} = -\frac{2}{3}$ (see Table 5).

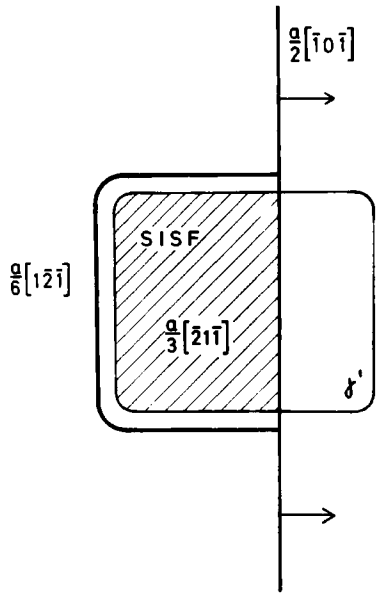


Figure 51. Schematic view of the creation of stacking faults inside a γ' precipitate.

because $(a/2)\langle 112\rangle$ partial dislocations have not been observed. The presence of an $(a/6)[1\bar{2}\bar{1}]$ loop together with $\boldsymbol{R} = (a/3)[\bar{2}1\bar{1}]$ or $(a/3)[112]$ leads to the following dislocation dissociation mechanism of a unit dislocation:

$$(a/2)[\bar{1}0\bar{1}] \rightarrow (a/3)[\bar{2}1\bar{1}] + (a/6)[1\bar{2}\bar{1}]$$

This particular dislocation reaction is confirmed by the presence of unit dislocations which were apparently obstructed at the γ–γ' interfaces. The situation is shown schematically in Fig. 51.

As a result of the $\langle 110\rangle$ grain orientation, two slip systems are easily activated. The occurrence of two slip systems also causes typical faults. Some of the stacking faults inside the γ' precipitates indeed reveal a line of intersection of two $\{111\}$ slip planes. By edge-on tilting it was found that part A of a fault such as that in Fig. 52 is located on (111) along $[\bar{2}11]$, and part B

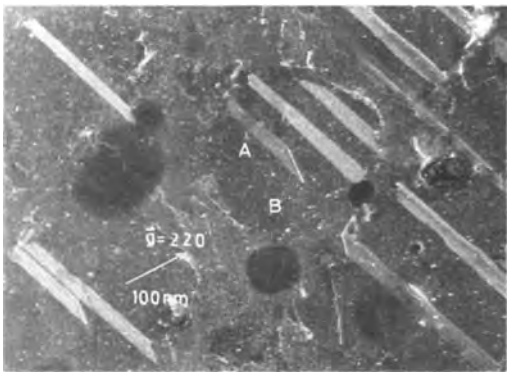


Figure 52. A stair rod dislocation is visible along the line of intersection of the two-fault plane.

on $(11\bar{1})$ along $[1\bar{2}\bar{1}]$. It is likely that the $(a/3)\langle 112\rangle$ partial dislocations, which have created the faults, have formed a stair rod dislocation at the line of intersection of the two slip planes. In fact, several stair rods can be formed if two $(a/3)\langle 112\rangle$ dislocations on these slip planes meet. Regarding the self energy, $(a/3)\langle 110\rangle$ is the most likely. Assuming the leading $(a/3)\langle 112\rangle$ partial dislocations which have created the fault to be of pure edge character, the exact Burgers vector may be deduced from the direction of the faults. This implies a dislocation reaction according to

$$(a/3)[\bar{2}11] + (a/3)[1\bar{2}\bar{1}] \rightarrow (a/3)[\bar{1}\bar{1}0].$$

In silicon, grown-in faults have been observed with a similar stair rod dislocation [137]. The stair rod in Fig. 52 is one side of a triangle which reveals a contrast typical for two faults, one above the other, which have a symmetrical inclination to the foil. The fringes are modulated parallel and perpendicular to the stair rod dislocation.

From these results it is concluded that upon deformation at temperatures near 800°C, $(a/3)\langle 112\rangle$ partial dislocations are

mobile in spite of a high Peierls stress. At room temperature, deformation until fracture produces a rather low density of twins. Calculations show that $(a/3)\langle 112\rangle$ partial dislocations may have glissile and sessile core structures. Therefore, the explanation might be that at room temperature $(a/3)\langle 112\rangle$ partial dislocations are less mobile [138]. From the separations between two $\frac{1}{2}\langle 110\rangle$ dislocations the APB energy is estimated to be $144 \pm 15\,\mathrm{mJ\,m^{-2}}$, which is less than the fault energy calculated (see Table 3) for pure Ni_3Al. Veyssière [139] has calculated that an $(a/3)\langle 112\rangle$ partial dislocation has a stable position at a certain distance from a Shockley partial dislocation provided $\gamma_{SISF}/\gamma_{APB} < 0.1$. In the present case this would imply $\gamma_{SISF} < 14\,\mathrm{mJ\,m^{-2}}$. However, a spontaneous dissociation of a $\langle 110\rangle$ dislocation into two $(a/3)\langle 112\rangle$ partial dislocations has not been observed. Therefore, $\gamma_{SISF}/\gamma_{APB} > 0.03$, that is, $\gamma_{SISF} > 4\,\mathrm{mJ\,m^{-2}}$.

At higher temperatures dislocation climb cannot be ignored. A typical example of a superlattice dislocation pair with a Burgers vector of $(a/2)[\bar{1}0\bar{1}]$ is depicted in Fig. 53. Two micrographs forming a stereo set were used to construct a three-dimensional image. A well-defined image of the dislocation has been obtained by a deviation from the Bragg position $w = 8.7$. To determine the value for $w = s_g \cdot \xi$ one needs to know the extinction distance ξ. Assuming a statistical distribution of the elements over the relevant sublattices in $L1_2$, ξ is found to be 42 nm. At first sight the superlattice dislocation can be divided into three parts [140]. However, a more detailed identification of the crystallographic planes revealed four sections, which are specified in

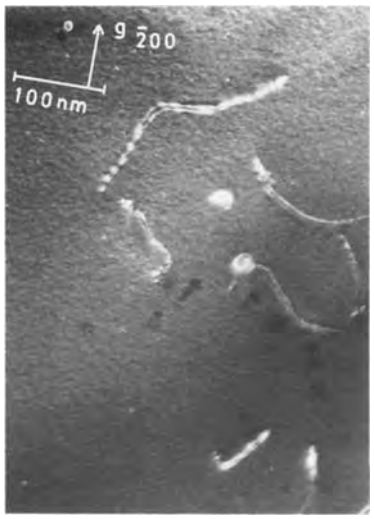


Figure 53. TEM image of a climbed superlattice dislocation.

Table 6 and schematically represented in Fig. 54. It may be surprising that only one segment of the superlattice dislocation is dissociated on a cubic plane whereas dissociation on an octahedral plane does not occur at all. The second, third and fourth segments have not reached their actual position by pure glide. It was also found by Veyssière et al. [141] in polycrystalline Ni_3Al that screw dislocation segments were dissociated on {001} planes whereas nonscrew segments were dissociated by a combined glide–climb process. The results listed in Table 6 confirm these findings. The fact that dissociation occurs by a thermally activated process may offer

Table 6. Specifications of Figs. 53 and 54 (length measured in the (001) projection).

Plane	Direction	Length (nm)
(010)	$[\bar{2}0\bar{1}]$	73
(112)	$[\bar{1}10]$	46
(101)	$[041]$	71
(110)	$[\bar{3}3\bar{4}]$	107

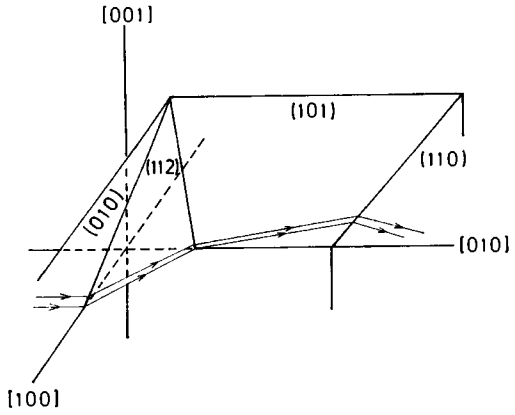


Figure 54. Schematic representation of the dissociation depicted in Fig. 53.

an explanation of the anomalous temperature dependence of the strength of $L1_2$ ordered alloys. At a certain temperature the dislocation segments on nonglissile planes may become mobile since the Peierls stress is lowered at increasing temperature. However, it should be emphasized that any detailed explanation of the yield stress anomaly in terms of dislocations must invoke intrinsic properties of dislocations, like core structure, mobilities and dissociations. At least six different models are proposed in the past to explain also related properties of the rate sensitivity and thermal reversibility [142–151].

At the end of this section it should be mentioned again that the tilting in the deformation holder was limited to a single tilt ($\pm 30°$), and thus the specimen sometimes could not be double tilted to the right with respect to the electron beam so as to excite the different reflections that are necessary in a complete characterization and analysis of the defects present in the sample. As a consequence, after the in situ deformation, specimens that promised to be interesting were shaped into 3 mm diameter discs and mounted in a TEM holder with two tilt axes. After the development of these in situ single-tilt TEM holders in the early 1970s [152–156], new double-tilt heating-straining stages became available (illustrations can be found in the literature [157, 158]).

Thus far we have been concerned with relating physical properties of metallic systems to structural information, either static or dynamic, but still on a micrometer scale. The question addressed in the following section is whether it is necessary and relevant to go beyond this scale, that is, to look for structural information on an atomic scale.

1.4.3 High-Resolution Transmission Electron Microscopy

Generally speaking, two ways of imaging in the transmission electron microscope can be distinguished. In conventional TEM, diffraction contrast is employed where the contrast in the image is determined by the intensity of the imaging beams, that is, the beams enclosed by the objective aperture. As is shown in the previous sections, rather low resolution information about the structure of the sample viewed along the direction of the incident beam can be obtained. In modern analytical transmission electron microscopes with reduced spherical aberration, a larger objective aperture can be used by which a large number of diffracted beams may interfere in the image improving the image resolution. With the advent of (ultra-)high resolution TEM (HREM) it is nowadays even possible to derive from the images structural information at an atomic scale. The technique of HREM

has its origins in the technique of phase contrast microscopy, which was introduced by Frits Zernike [159] of the University of Groningen for optical microscopy. In 1953 he received the Nobel prize in physics for this invention, which has found particular applications in the biological sciences.

HREM imaging is based on the same principles of phase contrast. Phase contrast imaging derives contrast from the phase differences among the different beams scattered by the specimen, causing addition and subtraction of amplitude from the forward-scattered beam. Components of the phase difference come from both the scattering process itself and the electron optics of the microscope. In a weak phase object the amplitude of the Bragg scattered beam is small compared to that of the forward-scattered beam. Most high-resolution images are made under conditions where visual contrast is best, that is, usually corresponding to a small defocus value.

The formula for the combined phase shift due to the optics of the microscope can be formulated at [160–162]

$$\chi(u) = \frac{2\pi}{\lambda}\left[C_s\left(\frac{\lambda^4 u^4}{4}\right) + \Delta f\left(\frac{\lambda^2 u^2}{2}\right)\right] \quad (19)$$

where C_s represents the spherical aberration coefficient of the instrument. The microscopist may control the amount of defocus Δf, and thus the phase shift of the off-axis beam. Underfocusing the lens ($\Delta f < 0$) compensates in part for the detrimental effects of spherical aberration. Overfocusing exaggerates the effect. Contrast originates from the difference in amplitude between the incident wave and the sum wave after scattering. The phase contrast transfer function $T(u)$ is related to $\chi(u)$ and the aperture function $A(u)$ by

$$T(u) = A(u)\sin\chi(u) \quad (20)$$

which describes how the microscope affects the phases of the imaging beams located at positions u (reciprocal coordinate). The problem, however, is that the contrast in fact oscillates, sometimes light against dark, and sometimes dark against light. Some amount of underfocus ($\Delta f < 0$) is needed to counteract the effects of spherical aberration. The optimum defocus condition was presented by Otto Scherzer [163], and is the value for which $|\sin\chi(u)|$ is close to unity for the largest possible range of u. The optimum defocus value is then found by setting the derivative of $\sin\chi(u)$ with respect to u equal to zero for $\chi = 2\pi/3$, leading to

$$\Delta f_s = -\sqrt{\tfrac{4}{3}C_s\lambda} \quad (21)$$

The $\sin\chi(u)$ crosses the axis for $\chi = 0$, that is (see Eq. (19)),

$$u_m = 2(3C_s\lambda^3)^{-1/4} \quad (22)$$

The resolution, defined as the least resolvable distance in the image, is then taken to be equal to u_m^{-1}. In that sense the resolution of HREM has traditionally been defined in terms of its Scherzer resolution limit at optimum defocus. Even beyond the Scherzer limit, spatial frequencies can be transferred from the specimen to the image, out to the so-called information limit of the microscope. The image detail thus obtained, however, is no longer directly interpretable due to the oscillatory contrast behavior. Coene et al. [164] have proposed a phase retrieval procedure to make the high-resolution information interpretable, which provides the aberration-corrected electron wave from a

focal series of HREM images. The ultimate goal of course is to achieve 0.1 nm structural resolution by using the information limit rather than the point-to-point resolution of the microscope. The information limit is extended to about the 0.1 nm range by the much better spatial and temporal coherence of the field emission gun as compared with the commonly used LaB_6 sources. Besides variation of focus as a method of phase retrieval, electron holography [165] seems to be a promising technique with which to achieve 0.1 nm structural resolution.

Roughly speaking, HREM of metallic systems can be divided into two different categories: one dealing with the atomic structure of defects and the other with chemical ordering in ordered compounds. The latter in particular comprises a panoply of structures. A rich literature exist on superstructures, domain structures and stacking disorder to which the Antwerp group under the leadership of Amelinckx and Japanese groups have contributed most substantially [165–175]. In itself it is understandable that ordered compounds are relatively popular here since these alloys are usually well ordered along the projection axis and, as shown by Amelinckx et al., it is feasible to obtain chemical and topological information by reference to either bright field or dark field images that may reveal the positions of the minority atomic columns [171]. More recently a detailed study of the atomic structure of antiphase boundaries in $L1_2$ ordered compounds has been published as well [176, 177].

In the category of defects the work on metals using HREM is less extensive. Work has been done on point defects in metals, related to radiation damage [178, 179], small metallic clusters [180, 181] and bulk dislocations [182–184], mostly in semiconductors [185–187]. However, following the CTEM work on interfaces, most HREM investigations have been devoted to interfaces for obvious reasons since many interesting physical properties are in fact determined by their atomic structures. However, in contrast to the work on ordered compounds, this does not mean this is an obvious choice from an HREM point of view. Quite frequently the boundaries in metallic systems are stepped or curved so that the requirements needed to determine the atomic arrangements are not met. In principle, the boundary should have an edge-on configuration to the incident beam direction with no interfacial steps normal to the beam direction. However, when the number of steps is still at a rather low density, high-resolution imaging might still be useful. Further, edge-on boundaries are most useful indeed, but HREM could still have some meaning in the study of inclined boundaries [188].

One of the earliest studies on grain boundary structures in aluminum and molybdenum were carried out by Bourret, Penisson and Gronsky [189, 190], at relatively low accelerating voltage (100 kV), whereas at high voltage (500 kV) tilt boundaries in gold were investigated by Ishida's group [191]. More examples are given by Smith and Barry [192].

At first sight one would think that all the aforementioned constraints would limit the usefulness and potentiality of HREM substantially to the field of materials science. So, the question to be asked again is whether HREM can help to solve materials problems, and not just reveal the structure of interfaces at some

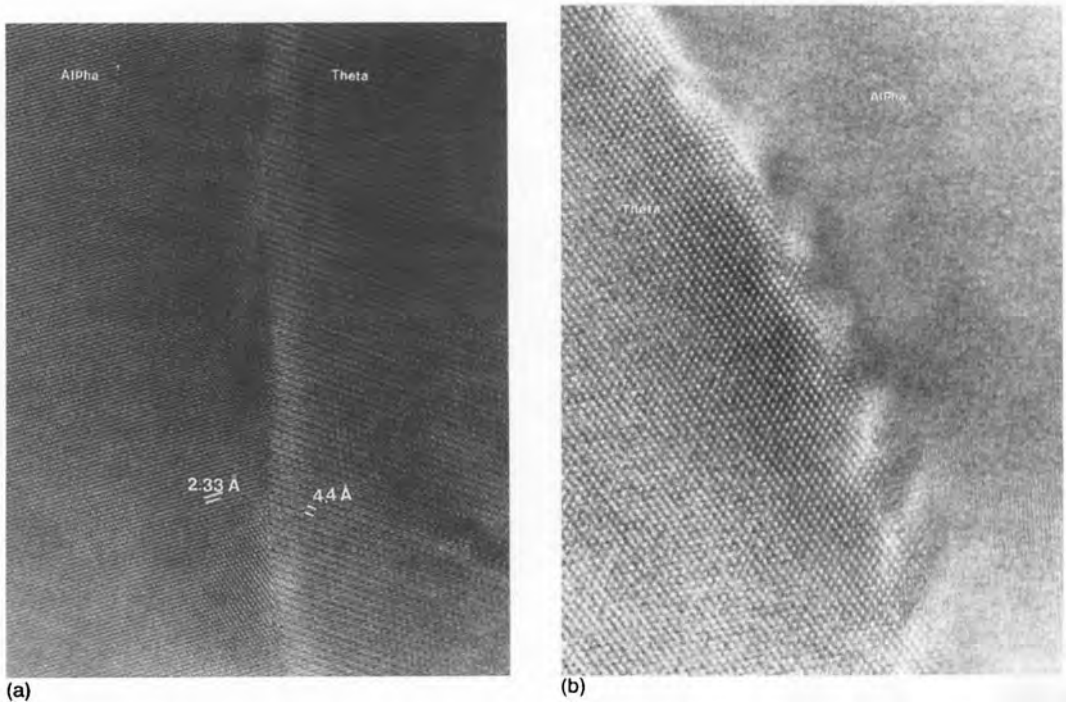


Figure 55. (a) α/θ interface in $[110]_\alpha$ zone axis orientation. Note the overlap between the phases. (b) α/θ interface. Note the ledge-type interface structure.

level of resolution. Fortunately, special boundaries dictated by symmetry arguments are more favored and relevant for some of the material properties than general boundaries (for a review, see Sutton and Balluffi [193]). In addition, as a matter of course, short periodic boundaries derived from symmetry arguments are much more popular with computational materials scientists, as a result of which HREM observations are often combined with atomistic calculations [194, 195]. This more recent alliance between experimental microscopists and theoreticians has turned out to be extremely useful and to some extent even essential.

To illustrate the possibilities and difficulties with HREM analysis, a typical example is shown in Fig. 55, where precipitation reactions are used to impose symmetry constraints for HREM research. To be somewhat in line with the previous sections, a θ interface is depicted inside an α matrix close to a [110] zone axis orientation in a laser-treated Al–5 wt.% Cu alloy (laser beam velocity $0.00125\,\mathrm{m\,s^{-1}}$). The magnification was calibrated on the 111 fringes in the α phase, which are $d_{111} = 0.233\,\mathrm{nm}$ apart. The fringe spacing inside the θ phase is 0.44 nm, which corresponds to the interplanar spacing of 110 planes. The lighter area at the interface clearly indicates that the two phases overlap there, and although there is contrast in both phases, the interface between them is not edge on. Figure 55b shows clear contrast in the θ phase, and little or no

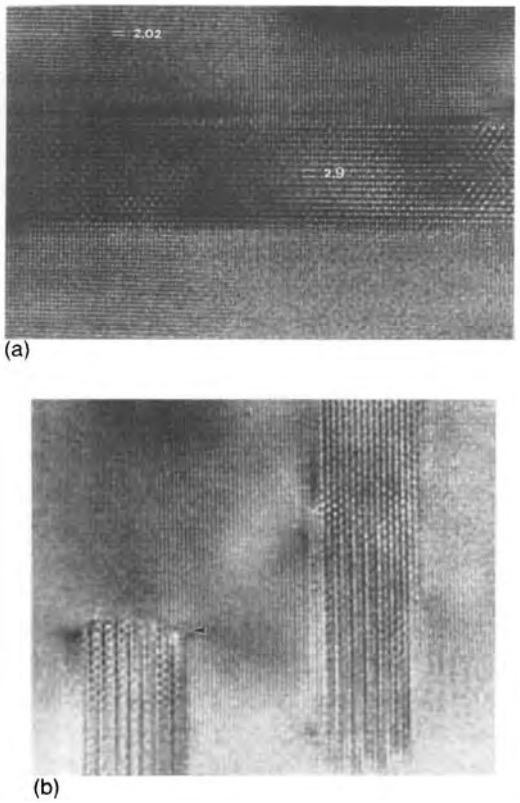


Figure 56. (a) θ' plate in [100] zone axis. (b) $[100]_{\theta'}$. The arrow indicates a misfit dislocation.

contrast in the α region. The wide parallel areas at the interface indicate the presence of some sort of ledge-type structure at the interface. Again this tells us that the interface is not edge on. So, this α/θ interface is not a good candidate for HREM. A much better choice is to focus on θ' precipitates which are semicoherent with the matrix. In Fig. 56 the matrix is tilted to a [100] orientation, showing $d_1 = 0.202\,\text{nm}$ fringes and illustrates how the (001) planes of the matrix are coherent with the (001) planes of the θ' precipitate. The fringes in the plate that are parallel to the interface were measured to be $d_2 = 0.29\,\text{nm}$ apart, which corresponds to the (002) fringes in θ'. Along the plate the contrast inside the plate changes from rectangular to diagonal fringes. Apparently the foil has a variation in thickness which causes the change in contrast. The thickness of the plate is estimated to be 17 atomic planes. At the end of the platelets, Moiré fringes can be seen in Fig. 56b. Their spacing is 0.66 nm, which corresponds to a Moiré fringe spacing expected from the overlap of the c axes of α and θ: $d_1 d_2/(d_1^2 + d_2^2 - 2d_1 d_2)^{0.5}$. At the top of the left-most plate an extra half plane can be seen in the matrix (arrow), which can be identified as a misfit dislocation accommodating the mismatch in lattice parameter. It has been shown [193] that θ' may nucleate on existing $\frac{1}{2}\langle 011\rangle$ dislocations that are dissociated by the reaction on a $\{001\}$ plane according to $\frac{1}{2}\langle 001\rangle + \frac{1}{2}\langle 010\rangle$. Therefore the interface can be regarded structurally as a stacking of partial $\frac{1}{2}\langle 001\rangle$ dislocations in the matrix that accommodate the mismatch with the precipitate. Similar symmetry-confined precipitates are found in the Al–Ag system [196, 197].

So far we have summarized the potential applications of HREM to structural aspects in metals and illustrated some of the complexities in resolving the structure at an atomic level. Actually, we have not referred to any material property that was elucidated by HREM itself. To be consistent within the scope of this contribution let us direct ourselves to a material property such as intergranular fracture and ask ourselves whether HREM has contributed towards solving the basic question in materials science why some additional elements in metallic systems promote embrittlement and others do not. In fact, there are not many clear examples presented in the literature, which is after

all mainly devoted to structural aspects as such and not directly to metallic properties. In contrast, a very nice example is presented by Luzzi et al. [198] studying the segregation of bismuth on grain boundaries in copper. In their synergistic approach of theory and experiment not only the structure of a copper grain boundary is characterized, as done before [194, 195], but also the chemical nature of the atoms is revealed. A long-standing materials problem here is to provide some clue why bismuth in particular leads to embrittlement of copper, while antimony, for example, does not. As both are in the same column of the periodic table, this would imply that the chemistry of these two elements should be about equal and the most important difference would be due to size effects. Ference and Balluffi have shown that segregation of bismuth to the grain boundary induces a faceting transformation at the boundary [199]. Although this is a quite common observation in this system, it turns out that $\Sigma 3$ boundaries in particular exhibit this behavior very strongly, that is, about 30% of the total grain boundary area is made of faceted $\Sigma 3$ boundaries. Luzzi et al. [198] were able to solve the atomic structure of the $\Sigma 3(111)//(11\,1)$ boundary by HREM. An example of their result is depicted in Fig. 57. The experimental image (left column) is compared with a calculated image at various defocus values. The images were obtained from grain boundary structures, calculated using *N*-body empirical potential as derived from the ab initio full-potential linear muffin tin orbital method, within the multislice formalism. The interesting point is that, due to the size effect of bismuth, the pure twin in copper with a stacking of A–B–C–B–A is

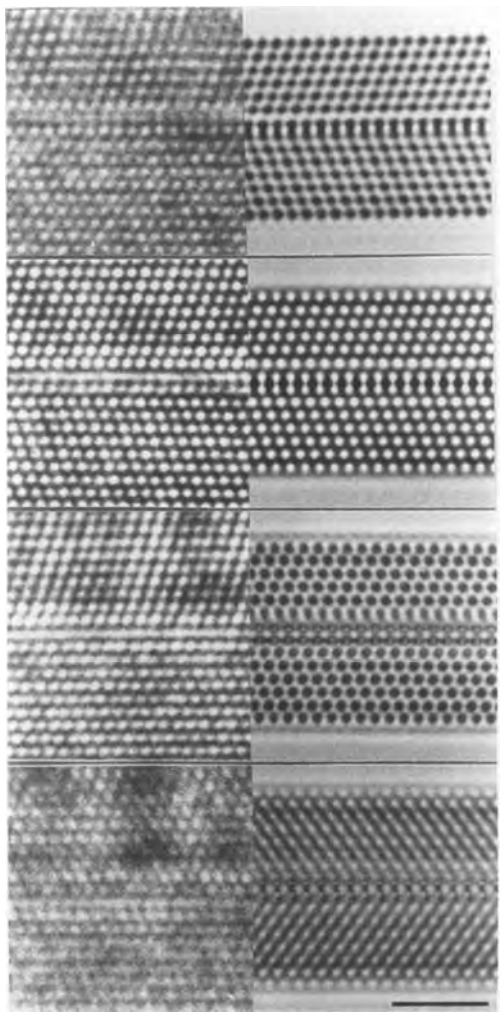


Figure 57. (a) Experimental (left column) and calculated (right column) images series with the objective lens underfocus increasing down the figure (−31.6, −63.6, −82.8 and −101.0 nm from defocus, respectively). The marker represents 1.5 nm. (Courtesy D. Luzzi)

altered to A–B–C′–C(Bi)–B–A. In fact, the bismuth atoms are centered outside the original C plane (Fig. 58). Although it is hard to say that we have a thorough understanding how embrittlement works in the Cu–Bi system (plasticity has been neglected) a comparison with antimony is

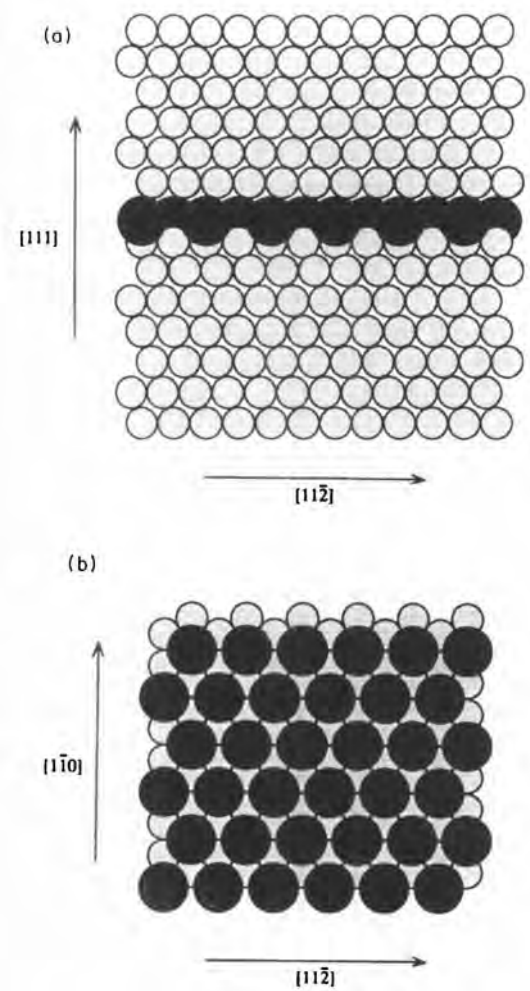


Figure 58. The relaxed final structure shown in two projections: (a) edge-on to the interface in the same direction as shown in Fig. 57b and in the plane of the interface. Bismuth atoms (dark) form an ordered hexagonal array with the interface between the adjacent copper atoms (light).

obvious now. Since the size effect of antimony is less, one would expect less faceted boundaries in the case of Cu–Sb in comparison with Cu–Bi and considerably less grain boundary expansion and a more ductile behavior instead. Indeed this is what is found experimentally [200]. Whereas faceted $\Sigma 3$ boundaries in Cu–Bi are as much as 30%, in Cu–Sb it is just 4%. Although the driving force for the (111)//(111) facets observed at the $\Sigma 3$ boundaries are also due to the pressure of the antimony, the large grain boundary expansion like in Cu–Bi system is not observed.

1.5 Imaging Metals and Alloys: Conclusions

The main emphasis of this contribution is on the structure–property relationship, so as to bridge the length-scale gap involved in the study of the stability of metallic systems. This section focuses on several aspects within this framework by looking at the detailed atomic structure on nanometer and micrometer scales, using TEM as the main tool. With all the possibilities that modern microscopes now allow [201], it is usually argued that it is necessary to characterize the microstructure at the highest levels of resolution possible in order to understand material behavior and to facilitate the design of either new or improved metallic systems. The question is whether this is really true. In other words: do we really solve materials problems, and what level of resolution is appropriate? Is the highest resolution the best route to understand properties which are structure-sensitive?

The basic philosophy should be that depending on the material property one wishes to investigate, it is necessary to consider the appropriate length-scale where the relevant processes take place. The challenging part of this rather mesoscopic approach is of course to define the

physical basis at the various length-scales and the corresponding transitions involved. For that reason, various case studies are presented here.

This section deals with metallic systems. We have not attempted to include every microscopic aspect of metallic systems. The important field of transformation in metals was only slightly touched upon. Reviews and applications of TEM in this field can be easily traced in the literature [202–205]. It is simply impossible to cover five decades of microscopic research on metals and alloys in one single review or even in one book. For that reason we took the different approach of explaining several case studies. Although the principal motivation for TEM of metals and alloys is to image microstructural defects and to derive from these observations their influences on the physical properties, the underlying scientific message is that one should be critical in the microscopic evaluation, for example, which defects are relevant, what kind of information is needed and, most importantly, on what scale?

The structure–property relationship of materials is in itself a truism. However, the actual linkage between structural aspects of defects in a material studied by TEM on one hand and its physical property is elusive. The reason for this is that many physical properties are actually determined by the collective behavior of defects rather than by the behavior of one singular defect itself. For instance, there exists a vast number of TEM analyses in the literature on ex situ deformed metals which try to link observed dislocation patterns to the mechanical behavior characterized by stress–strain curves. However, in spite of the enormous effort that has been put into both theoretical and experimental work, a clear physical picture that could predict the stress–strain curve on the basis of these TEM observations is still lacking.

There are at least two facts that hamper a straightforward correlation between microscopic structural information and materials properties: one fundamental and one practical.

Of course, it has been realized for a long time that in the field of dislocations and interfaces we are facing highly nonlinear and nonequilibrium effects. The defects determining many physical properties are in fact not in thermodynamic equilibrium and their behavior is very much nonlinear. This is a fundamental problem since an adequate physical and mathematical basis for a sound analysis of these highly nonlinear and nonequilibrium effects does not exist. Nevertheless, the situation is not hopeless since there are two approaches to circumvent these problems, and TEM may still contribute quite a lot. One had to do with numerical simulations that incorporate the behavior of individual defects as known from both classical theory and from TEM observations of individual dislocations/interfaces [206–208]. For example, to have a thorough understanding of the generation of cellular dislocation structures, vein structures, tangles, subgrain boundaries and persistent slip bands, important input for these numerical simulations on the behavior of individual dislocations such as cross-slip behavior, climb and bipolar structures, and so forth, may be provided by TEM research. The other approach is the counterpart of numeric simulations, namely a continuum mechanics approach, initiated by Holt [7]. This provides a description of

the global cooperative behavior of defects [209–216] and focuses on the instability transitions and accompanying structural transformations. Here also, experimental knowledge provided by TEM, in combination with complementary techniques [217], is inevitable.

To correlate the microstructural information obtained by TEM of an interface to the macroscopic behavior of polycrystalline solids is even more complex than in the case of dislocations. The reasons for this are numerous, for example the limited knowledge of the interface structure, that is, both topological and chemical, at an atomic level of only a small number of special cases, the complexity due to the eight degrees of freedom of an interface and the lack of mathematical–physical models to transfer information learned from bicrystals to the actual polycrystalline form. It has been shown that in some cases it is crucial to have information on an atomic level available but surely it is not always necessary and sometimes rather more appropriate to image defects on a micrometer scale instead to correlate the structural information to physical properties. An interesting approach to this end, that is, to couple boundary to boundary variations and to couple processes in the boundaries to the grain interiors, is presented by Nichols et al. [218–220] based on the concepts of percolation theory. Actually, these workers introduce new measures at a different length-scale and microscopic features such as the radii of gyration and clusters that are joined by like boundaries. Therefore, it is argued that for a more quantitative evaluation of the structure–property relationship of materials it requires a de-emphasis of analysis on an atomic scale [221].

Another more practical reason why a quantitative TEM evaluation of the structure–property relationship is hampered has to do with statistics. This is not specific of course for metallic systems. The metrological considerations of quantitative TEM from crystalline materials have been summarized by Goodhew [222, 223], one of which is the statistical significance of the TEM observations. In particular, in situations where there is only a small volume fraction of defects present (see Sec. 1.4.1 of this Chapter) or a very inhomogeneous distribution, statistical sampling may be a problem. An interesting plot is produced by Chone [224], depicted in Fig. 59. It indicates the necessary volume to be examined as a function of volume fraction of particles of various sizes with a precision of about 20%. So, in our case of gas bubbles of the order of 10 nm and a volume fraction of the order

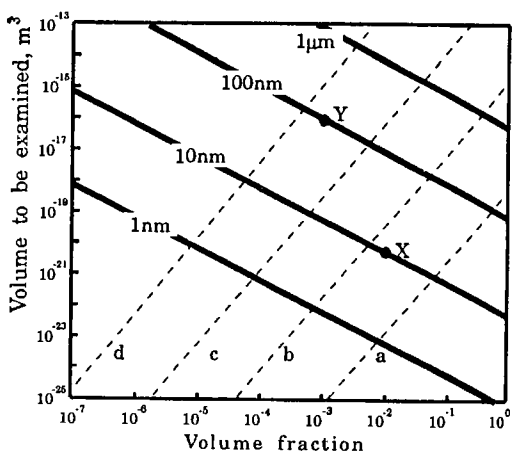


Figure 59. The volume to be examined as a function of the volume fraction at various diameters of inclusions (solid line). The dashed line refers to the number of fields of view (or screens) needed. ($a = 1$, $b = 10$, $c = 100$ and $d = 1000$.)

of 1% it is necessary to view about 10^{-20} m^3 (point X). At smaller sizes and volume fraction of course a smaller volume has to be examined (point Y). Goodhew has added to Fig. 59 lines of the particular number of fields of view (or screens) necessary, assuming that each field is explored at the most suitable magnification of a foil of uniform thickness of 100 nm. Here it would imply that for X about 10 screens are necessary, whereas for the situation at Y about 1000 micrographs are needed for an appropriate statistical sampling. In our case where bubbles are overlapping (see Fig. 4), quantitative measurements are almost impossible without applying stereo TEM (see Fig. 1). In the case of dislocations and interfaces one should also be aware of these aspects of quantitative metrology with TEM. For example, to support or reject various models of the structure–property relationship, such as the relationship between impurity segregation and interface structure, it is necessary to measure the boundary misorientation at very high precision [225], and they should all be aligned parallel to the electron beam for an effective quantitative analysis with TEM. This is virtually impossible, and only qualitative conclusions can be drawn.

Based on these fundamental and practical complexities of correlating quantitative microstructural TEM information to material properties, we conclude that TEM characterization, accompanied by complementary techniques, may serve as a crucial input for a deeper understanding of the structure–property relationship, but that one should focus more on the generic features of defects using a mesoscopic approach including various length-scale transitions.

Acknowledgements

The work described in this section is part of the research program of the Foundation Fundamental Research on Matter (FOM-Utrecht), and has been made possible by financial support from the Netherlands Organization for Scientific Research (NWO-The Hague). It has also been supported by the Netherlands Foundation for Technical Sciences (STW) and IOP-Metals (The Hague). Special thanks are due to Dr. Paul Bronsveld for scientific discussions and contributions. Thanks are also due to an exciting Ph.D. student–thesis supervisor relationship—equally important as the structure–property relationship!—in this field of electron microscopy on metals that I experienced over the years with Bert Huis in 't Veld, Hans de Beurs, Herman Hegge, Eric Gerritsen, Joeke Noordhuis, Bart Pestman, Willem-Pier Vellinga, Lodewijk de Mol van Otterloo and Paul Pos. Some of our joint work has been included in this contribution.

1.6 Imaging Phase Transformations in Metals and Alloys

1.6.1 General Introduction

Since the early days of the study of phase transformations in metals and alloys, microscopy techniques have played a crucial role in characterizing and understanding the features of the various phases and microstructures involved. Although this Handbook contains a separate chapter on phase transformations (see Chap. 2.2 of this Volume), the combination of both issues in one section allows us to present some previously undiscussed but typical microscopy contributions. For the basics of the different techniques mentioned, the reader is referred to the relevant chapters in the Volumes I and II of this Handbook. Some basic and recent works

on phase transformations, including metals and alloys, are listed [226–229].

From the viewpoint of experiments involving microscopy techniques, phase transformations in metals or alloys can readily be subdivided into different classes depending on the atomic rearrangements occurring during the transition (see also Chap. 2.2 of this Volume). The first class contains those transformations in which the basic lattice virtually remains unchanged but the actual occupation of a given lattice site will change: these transformations are called *diffusive* because of the relatively large distances (i.e., several interatomic distances) that atoms can travel during the transformation. In principle, all disorder–order transformations fall into this category. Due to the ordering, slight changes in the lattice parameters or symmetry of the basic lattice can occur, sometimes leading to specific microstructures [230, 231]. A typical example of this are the diamond-like plate configurations in CuAu thin films when transformating from the high-temperature disordered f.c.c. state to a low-temperature ordered orthorhombic structure [230]. The second class contains *displacive* transformations in which the main characteristic of the transition is the change in basis lattice between the parent and product phases. Here, there is no diffusion involved and thus the existing ordering is inherited through the transformation. The classic example of this type of transformation are the martensitic transformations [232] but the ω transition consisting of pure shuffles not changing the lattice parameters of the unit cell also belongs in this class [233]. Of course there exist a number of intermediate cases in which the transition has characteristics of both of the above extremes. Bainitic [234] and near ω-type transformations such as Ni_2Al (see below) qualify for inclusion in this intermediate class.

In the present section a few typical case studies each belonging to one of the above-mentioned classes will be discussed on the basis of the microscopic investigations performed on them. Examples of in situ experiments will be shown as well as cases in which information on the phase transformation was obtained through pre-treated samples. The former usually yields direct information on the transformation path while the latter allows for a higher resolution. For the diffusional transformations, some TEM studies of long- as well as short-range effects of disorder–order transformations in some binary alloys will be discussed. Contributions on different aspects of martensitic transformations including precursors, nucleation and growth obtained using different microscopy techniques such as optical microscopy (OM), scanning electron microscopy (SEM) and TEM are then reviewed. Finally, some examples in the field of mixed cases are also discussed.

1.6.2 Diffusive Phase Transformations in Alloys

Diffusive phase transformations generally introduce a change in symmetry, which results in domain formation due to the presence of different variants. These variants can be orientation variants, creating twin boundaries or inversion boundaries, or they can be translation variants creating APBs. A prediction of the number of orientation and translation

variants can be done based on simple group theoretical considerations [235–237]. In the present contribution we will concentrate on the microstructure resulting from ordering as well as on the microstructural changes when approaching a diffusion-controlled phase transformation. We will investigate the short-range order (SRO) when approaching the transition from above and the wetting of ordering defects such as APBs when approaching the transition from below.

We know that in general HREM images vary substantially with focus and sample thickness; however, for alloys, based on a simple lattice such as the f.c.c. (or b.c.c.) lattice, the interpretation simplifies considerably. The reason being that when such structures are viewed along a simple zone axis, for example [001], we obtain a column structure with the different columns separated by 1.8–2.0 Å. Such atom columns, parallel to the electron beam, constitute a potential well with cylindrical symmetry, and the electrons 'channel' through this potential well, continuously focusing and defocusing with a periodicity varying between 40 and 100 Å, depending on the nature of the constituent atoms (Fig. 60). One can show that as long as the columns do not come too close, electrons channeling in different columns will not mutually influence each other [238, 239]. Therefore for ordered alloys the image contrast of a particular column will only be determined by the chemical composition of that column and will not be influenced by the composition of neighboring columns [238]. This aspect of HREM imaging in column-type structures allows one to interpret ordering phenomena in ordered compounds.

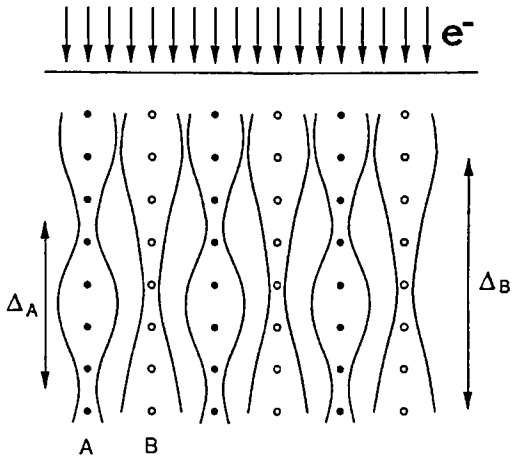
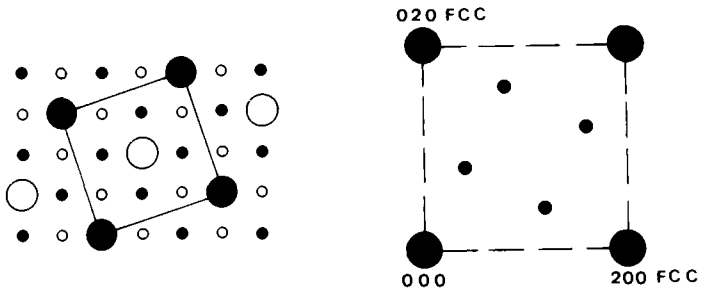


Figure 60. Electrons traveling through a column-type structure.

In practice, the choice of objective aperture used to select the diffracted beams contributing to the HREM image will determine the type of information which is observed in the image [240]. In Fig. 61 the different choices are shown for a [001] zone of the ordered Au_4Mn compound [241]. The unbroken circle is a dark field superlattice mode in which only ordering reflections are included after centering the aperture around the optical axis (i.e., under tilted illumination). In this mode the minority atom columns will always show up bright, and the image will be practically independent of thickness and defocus. An example of such an image is shown in Fig. 62 for a long-period superstructure in $Au_{11}Mn_3$ [242] where only the minority atom columns are visible. The higher intensity of the row of dots at the APB is also visible in simulated images and is due to the fact that the imaging mode is asymmetric with respect to the twofold axis of this boundary [240]. When the bright field basic reflections mode is used (dotted circle in Fig. 61),



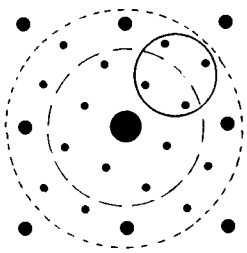


Figure 61. Different modes for HREM imaging by using different diffracted beams to form the image (Au_4Mn [001] pattern): unbroken circle, dark field superlattice mode, using only super-reflections; dashed circle, bright field superlattice mode; dotted circle, bright field basic reflection mode using basic as well as superstructure reflections.

the basic lattice will also become visible, and the actual image contrast will depend strongly on the imaging conditions to the point that it can show more bright dots than there are minority atom columns. An example of this is shown in Fig. 63 for Pt_8Ti, where the experimental image corresponds to the case of white dots on the positions of minority atoms. A slight change in thickness or defocus, however, would produce a white dot square with lattice parameters twice as small, as seen from the simulations [243]. The

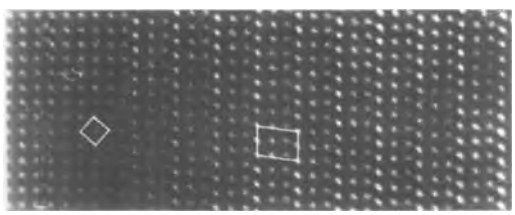


Figure 62. HREM example of long-period structure in $Au_{11}Mn_3$ using the dark field superlattice mode.

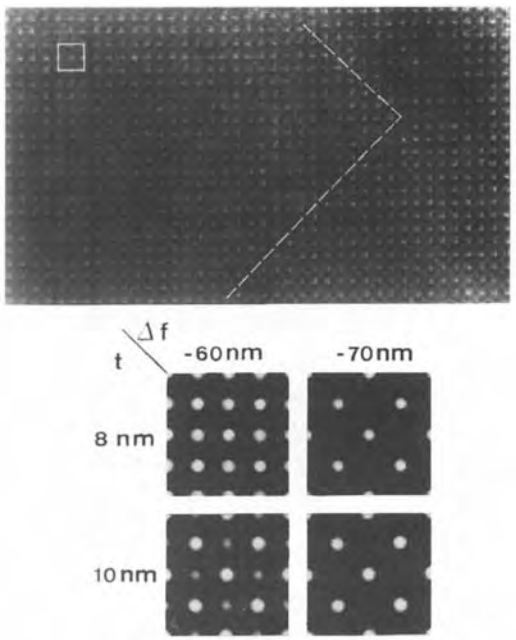


Figure 63. HREM example of Pt_8Ti obtained in the bright field basic reflection mode including simulations showing different white dot patterns for different conditions.

straightforward nature of the interpretation of these superlattice images in which only minority atom columns are made visible under known imaging conditions allows one to easily recognize and interpret new types of ordered structures or long-period variants of a given basic unity. Typical examples of this procedure are found in, for example, the Pt–Ti [244] and Au–Mn [241, 242, 245] systems.

1.6.2.1 Long-Range Ordered Alloys

It is clear that ordering in alloys can be detected by electron microscopy. Because of the lowering in orientation symmetry, twin variants or inversion domains will appear; if the translation symmetry is lowered, out of phase domains will appear. These different variants can be discerned even by conventional microscopy, and have been reported in the early literature on electron microscopy in materials science. A compilation of the different contrast effects produced by the various interfaces is shown in Fig. 64. More details on the identification of the different planar defects can be found in the literature [246]. With the recent developments of HREM it is possible to image the atomic configuration of such defects directly; we will treat two examples: Au_4Mn and $Cu_{3\pm x}Pd$. We will use these examples further when studying of the evolution of microstructure or the behavior of the alloy in the vicinity of the ordering temperature.

Au_4Mn as well as Ni_4Mo, Au_4V, Au_4Cr and Au_4Ti all have the D1a structure as the ordered structure, the structural representation of which is shown in Fig. 65c. As we will see later, all of them also have a SRO state of the type '$1\frac{1}{2}0$'. The disordered f.c.c. lattice has cubic symmetry, while the ordered structure has tetragonal $4/m$ symmetry. This results in six orientation variants; the tetragonal c axis can be along any of the cube axes and for every choice of the c axis, two orientations are possible of the a and b axes (see Fig. 65c). These six orientation variants, however, only produce two essentially different twin interfaces [241, 247]. The volume of the primitive unit cell of the ordered structure is five times larger than the primitive cell of the disordered structure, and therefore five translation variants, separated by four different APBs, result [235]. Ordering introduces a fragmentation into these different variants, resulting in patterns such as Fig. 66, where a twin interface between two orientation variants with a parallel c axis and two APBs are visible (compare with Fig. 65c). In Fig. 66 the Au_4Mn superstructure has been imaged using the dark field superlattice mode and only the manganese configuration is visualized as bright dots. The defects in the long-range order (LRO) state are now adequately studied and well understood [241].

The structure of Cu_3Pd is based on the $L1_2$ (Cu_3Au-type) structure. However, with deviating compositions from the 3:1 composition or with changing temperature, different long-period superstructures (LPSs) are found. Such LPSs consist of a regular occurrence of APBs, along the cube planes; depending on the displacement vector or the APBs, such APBs can be conservative or nonconservative. The average distance between the boundaries is termed M. At low temperatures and for well-annealed materials, the value of M is mostly commensurate and locked in at simple values (between 3 and 10 for $Cu_{3\pm x}Pd$ alloys) [248, 249]. An

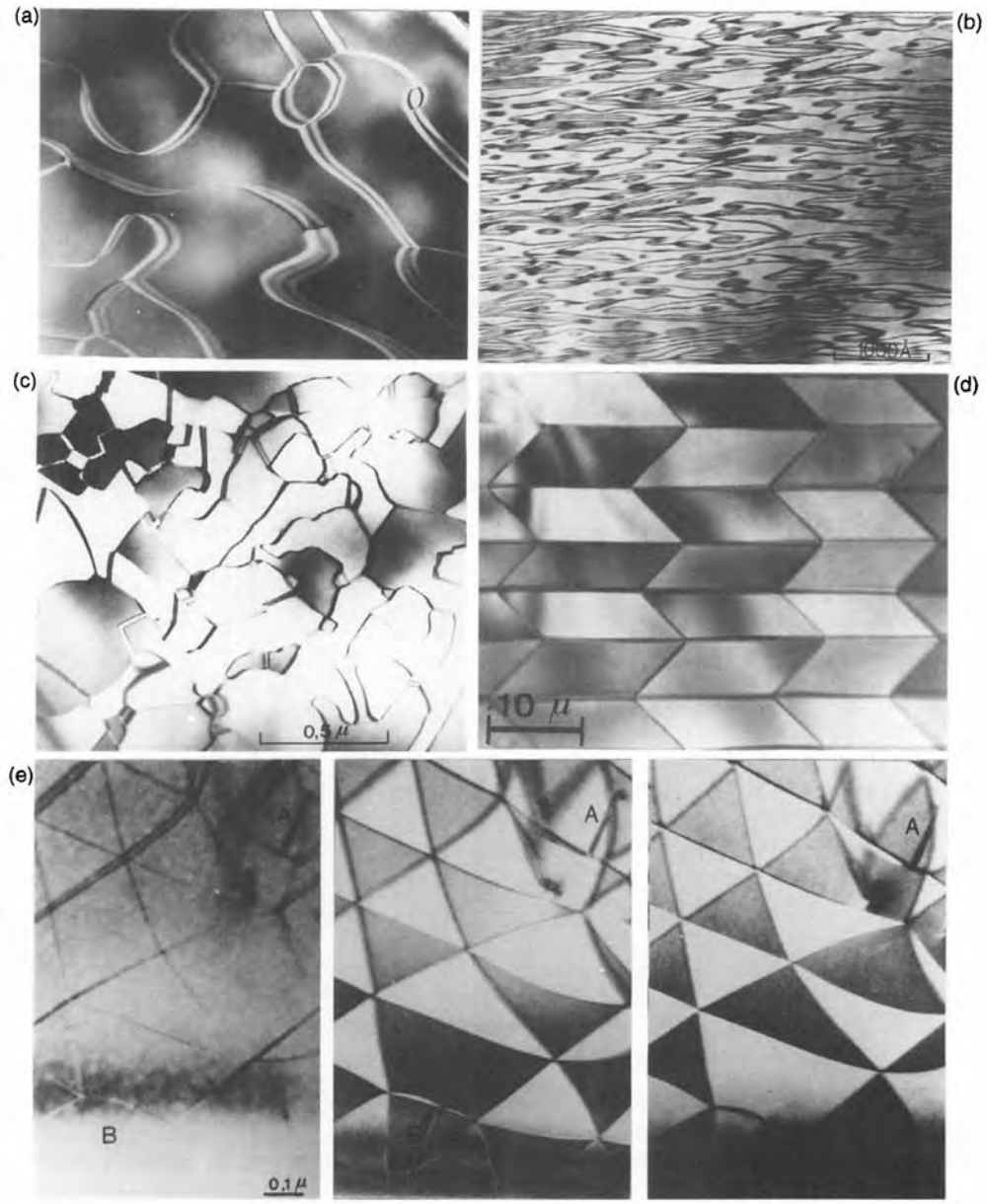


Figure 64. Diffraction contrast of different planar defects in long-range ordered materials. (a) APBs in Ni_4Mo; no background contrast difference is observed on both sides of the interface. (b) High-density and pseudoperiodic arrangement of APBs in Ni_3Mo. (c) Twin domains with a common c axis in Ni_4Mo; the corresponding diffraction is Fig. 65b, and the schematic configuration is reproduced in Fig. 65c. Note the differences in background contrast between different twin domains. (d) Regular twin configuration in γ-brass (courtesy of M. Van Sande). (e) A bright field image and two dark field images of inversion boundaries in the χ phase (courtesy of M. Snykers). Inversion boundaries appear in APBs in bright field, but are characterized by contrast differences in multiple beam dark field images.

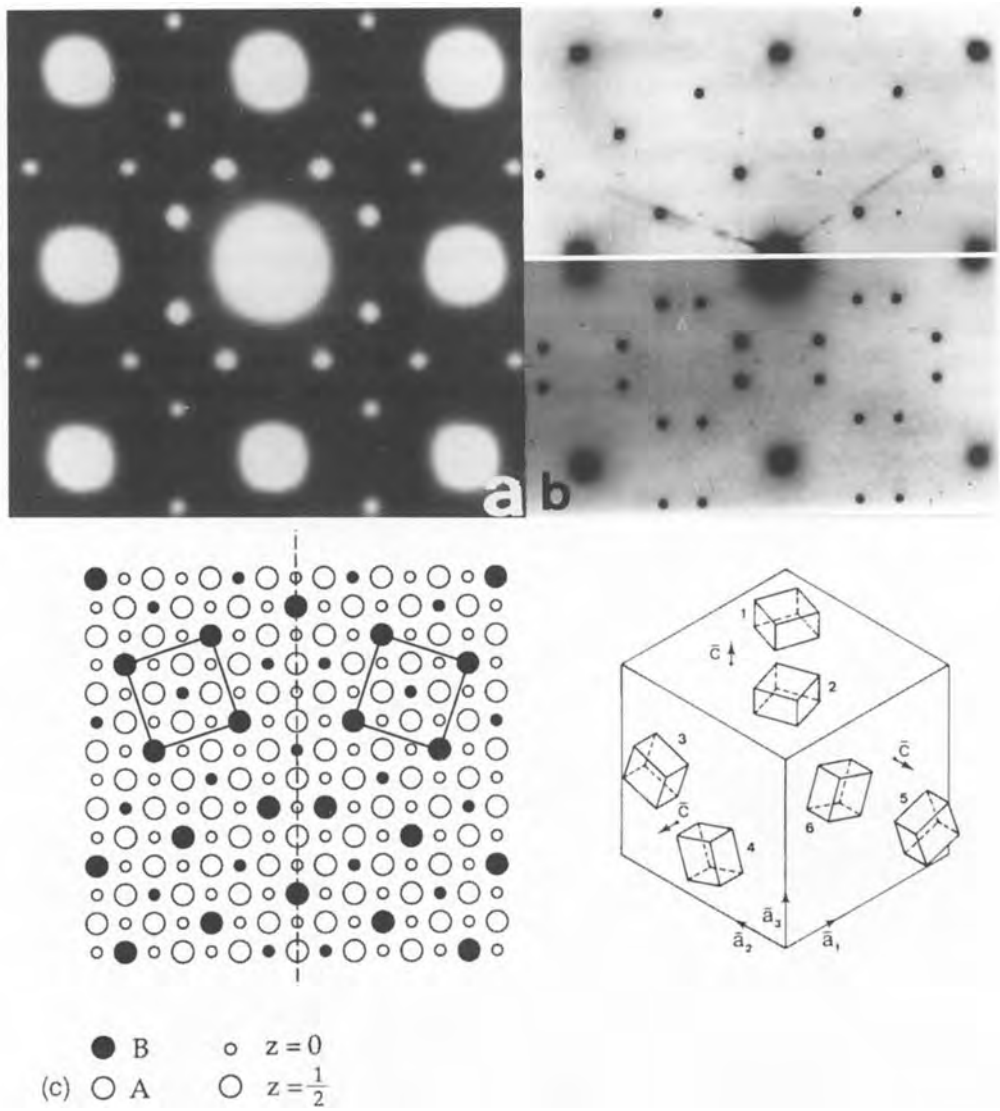


Figure 65. (a) [001] diffraction pattern of the '$1\frac{1}{2}0$' SRO state in Ni_4Mo. (b) [001] diffraction pattern of the LRO state in Ni_4Mo; the upper part is for one variant; the lower part for two variants with a common c axis. (c) [001] projection of the LRO state of D1a, showing two different orientation variants with parallel c axes and projection of the six different orientation variants on a cube.

image at high resolution is shown in Fig. 67. The interfaces are straight, sharp and well defined. Under the imaging conditions used, the bright dots can be identified as the palladium atoms; the shift at the individual interfaces is clearly visible.

1.6.2.2 Short-Range Order and Initial Stages of Ordering in D1a-Type Alloys

Studying SRO in alloys by electron microscopy is barely possible in situ since oxidation inside the electron microscope

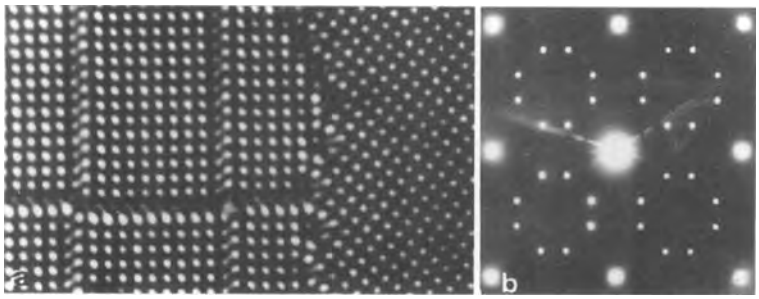


Figure 66. (a) High-resolution image of Au_4Mn with limited resolution. (b) The diffraction pattern, indicating the selected beams.

tends to limit the studies, even for alloys with a low transformation temperature. The only way to obtain information on the SRO state is by quenching the alloy from high temperature to room temperature and hoping that the state of order is retained. For a number of alloys, such as CuAu or CuPt, this procedure is not effective, ordering proceeds too fast and only the initial stages of LRO can be studied. For other alloys, however, the ordering is sluggish enough to maintain the SRO or the very initial state of LRO after quenching. Such kinds of alloy are chosen here to study the order–disorder transition. We have particular interest in so-called '$1\frac{1}{2}0$' alloys since the LRO reflections (mostly of the D1a-type structure) do not coincide with the SRO reflections (see Fig. 65). Examples of such alloys are Ni_4Mo, Au_4V, Au_4Cr, Au_4Fe, Au_4Mn and Au_4Ti. The SRO and LRO (e.g., in Ni_4Mo) has been studied since 1965 [247, 250], but discussion is still going on about the actual SRO state in a real space description: do we have ordered clusters, microdomains or a statistical distribution described by a concentration wave model? The actual difference between a cluster and a microdomain model is a matter of correlation length. The concentration wave model, first described by Khachaturyan [251], is basically a description in reciprocal space. The microdomain model and the concentration wave model, however, have to be considered as two descriptions of the same physical reality [252]; indeed, the materialization of a concentration wave in a discrete medium results in a geometrical pattern that can be considered as a cluster. In the past, a number of alloys showing a particular SRO or transition state between SRO and LRO have been studied in reciprocal space. The locus of diffuse intensity enables us to reduce a relation between some of the occupation sites of the fundamental lattice [253, 254]. Neutron and X-ray diffraction provide more accurate numerical data, but the fundamental limitation to pair correlations and the practical limitation to large selected areas still holds.

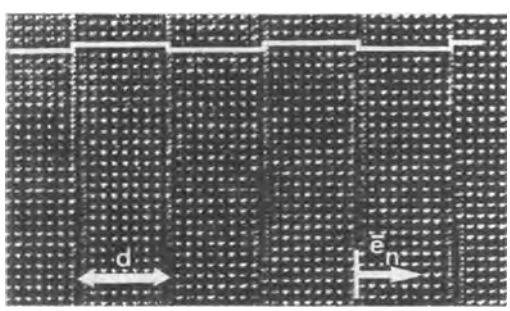


Figure 67. HREM of a commensurate LPS structure of $Cu_{3+x}Pd$ where APBs occur with periodicity $M = 8.0$.

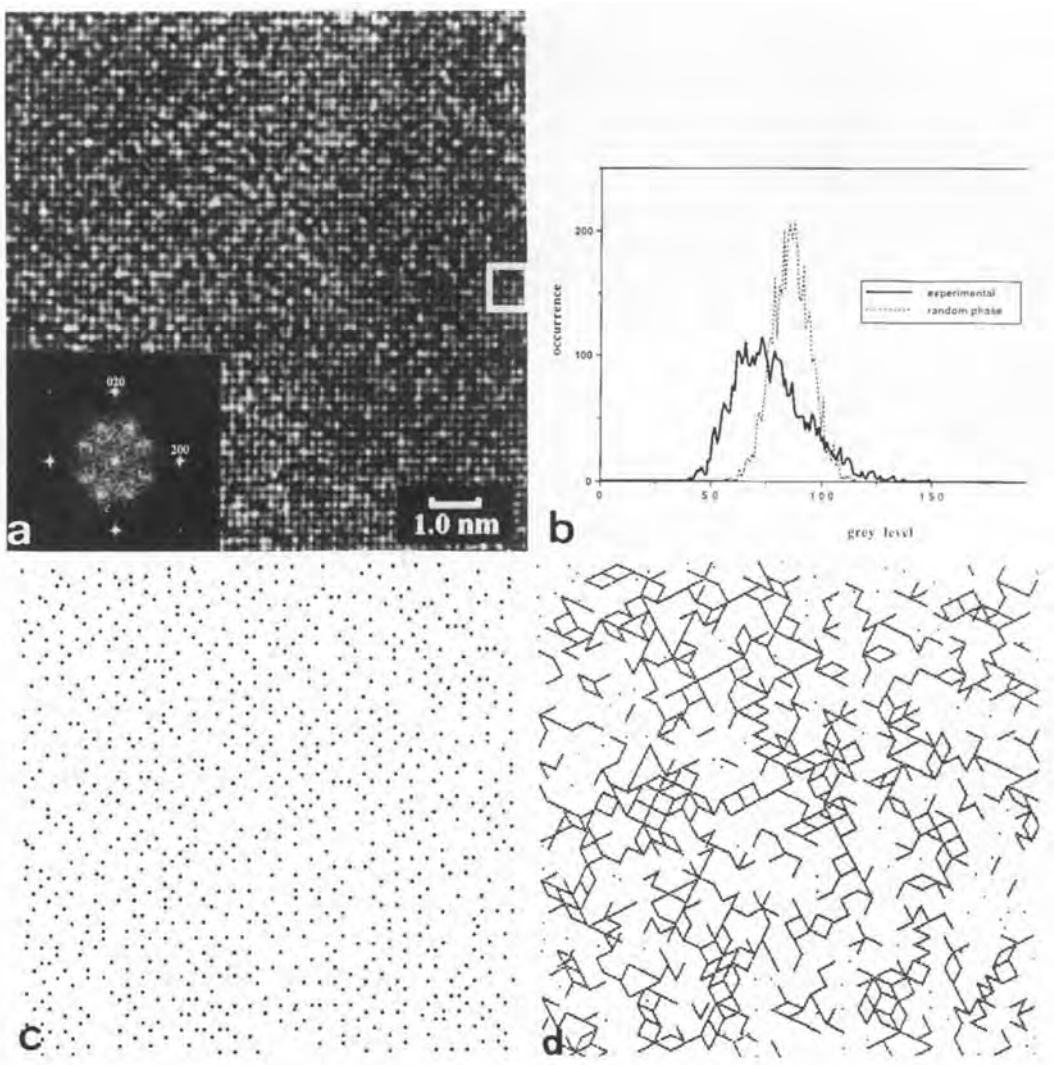


Figure 68. (a) HREM image of the SRO state in Au_4Cr; the power spectrum of (a), indicating the presence of SRO in the alloy, is shown as an inset. (b) A histogram of gray levels obtained from the SRO state, compared with a histogram from a 'random phase' distribution. (c) Result of image processing the original image (a). (d) The same as (c), but typical clusters of the type $\{2, 1\}$ have been indicated.

It is possible, however, to perform HREM observations on the SRO state in Ni_4Mo, Au_4Cr and Au_4Mn (Fig. 68). In spite of the fact that we have a column-type structure, based on a simple f.c.c. lattice, we realize that the interpretation of such images is far from straightforward. However, we have been able, with the necessary care and adapted image processing, to extract very useful information about the ordering. The Fourier transform of the HREM image of Fig. 68 indeed shows diffuse intensity similar to the electron diffraction pattern of Fig. 65a,

and therefore the ordering information is concealed within the HREM image. The main problem in HREM of such imperfect structures is the projection of a three-dimensional structure into a two-dimensional image. One can show, however, that for atom columns consisting of a random or almost random composition, the projection approximation still holds. Although this does not lead us to a direct correspondence between the column composition and its gray value—because of differences of thickness and focus—the differences in composition are still reflected in the image. If we classify for a disordered A_4B_1 alloy all [001]-type columns according to their chemical content, a binomial distribution is expected, which will be asymmetric [255]. For larger thicknesses, all columns will become equivalent, and the distribution will tend towards symmetry. We have digitized the experimental image of Fig. 68a and obtained a histogram of all gray levels (Fig. 68c); the asymmetry is clearly present. We can further show that whenever an asymmetrical histogram is obtained, the brightest dots in the HREM image correspond to columns deviating strongly from the average composition [255]. For images (such as Fig. 65a) which fulfill the binomial criterion, most information is found in those columns which appear less frequently, that is, which deviate most from the average composition. We can develop an image-processing routine to enhance the 'in-plane' correlations between these columns which give rise to the SRO intensity. From the experimental image of Fig. 68a we obtain a processed image (Fig. 68d). The Warren–Cowley SRO parameters $a_{i,j}$ can be directly calculated from the power spectrum, but also from counting the pairs in the binary image. It then turns out that pairs of the type $\{2, 1\}$ occur most frequently, indicating a preference of the structure to form identical columns separated by a vector $\{2, 1\}$. When these pairs are connected in Fig. 68d, we obtain the final image of Fig. 68e, where not only isolated lines can be discerned but also clusters of the D1a-type structure as well as of the DO_{22} structure. This is hardly surprising because the $\{2, 1\}$ pairs are the constituent vectors of both of these structures. We can now apply the same technique to the different stages of ordering between the SRO and the LRO states and follow the ordering kinetics. The Au_4Cr alloy is particularly suitable for such experiments because the ordering kinetics are very slow and therefore every state of order can easily be quenched in. Of importance—and beyond the scope of classic diffraction experiments—is the direct observations of clusters, larger than pairs, in these '$1\frac{1}{2}0$'-type alloys. We are aware that we have to include a certain 'error bar' for a quantitative interpretation of our results, because of ambiguities in judging the image intensities, but we are convinced that we are very near to the true projected structure. A strong argument being the fact that the obtained distributions are consistent with the experimental electron diffraction patterns. We therefore strongly believe that the present method opens new perspectives for studying SRO and ordering phenomena, including kinetics, in real space.

1.6.2.3 Interface Wetting in Ordered Alloys

When approaching the order–disorder transition from below, disordering will

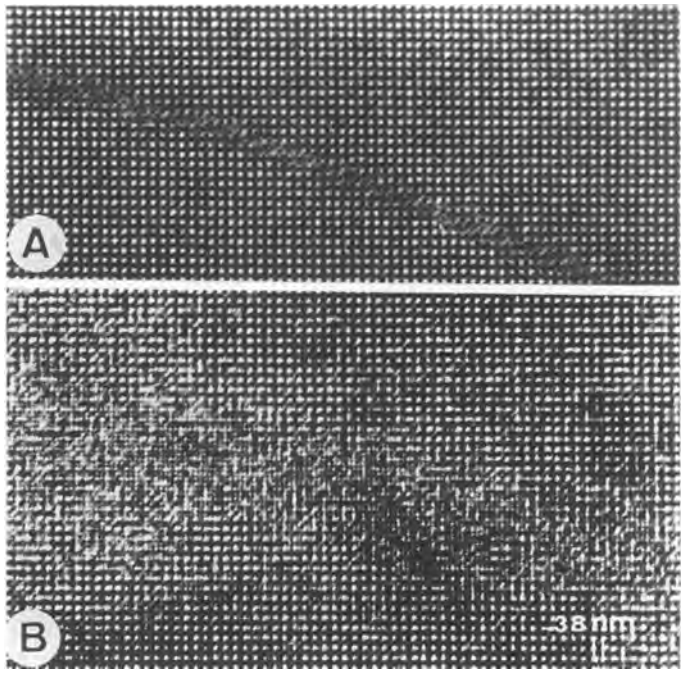


Figure 69. High resolution image along [001] of APBs in the $Pt_{70}Cu_{30}$ alloy (A) when annealed well below T_c and (B) when annealed just below T_c.

start from imperfections such as APBs [256] or twin interfaces [257]. Planar interfaces will be 'wetted' by the disordered phase at temperatures close to the transition temperature. We will first study the Pt–Co compound with the composition $Pt_{70}Co_{30}$, where the ordered structure is the $L1_2$. Further, we will consider the Cu–Pd compound which shows one- and two-dimensional long-period superstructures.

In the ordered compound $Pt_{70}Co_{30}$, APBs with a displacement vector of the type $\boldsymbol{R} = \frac{1}{2}\langle 110\rangle$ appear as a consequence of the decrease in translation symmetry. When the material is heat treated at temperatures well below T_c, these APBs show a normal behavior, that is, under two-beam conditions (0, $\boldsymbol{g}$), the boundary is out of contrast for $\boldsymbol{g} \cdot \boldsymbol{R} = n$. Under high-resolution conditions as in Fig. 69a the displacement vector can immediately be identified. The HREM imaging conditions are such that the bright dots in Fig. 69 represent the cobalt configuration. It should be noted that the structure is perfect up to the boundary plane and that the width of the interface is restricted to a single atomic plane. When annealing the material at temperatures close to T_c, the width of the interface gradually increases [258]; this is obvious from Fig. 69b. The APB is 2–4 nm wide, and at the interface the basic square f.c.c. lattice of 0.19×0.19 nm can be recognized, indicating the presence of disordered material between the two ordered regions.

In the Cu–17% Pd compound, Ricolleau et al. performed highly controlled heat treatments in the range between T_c −10°C and T_c with temperature steps of 0.2°C in order to characterize the nature of the divergence of the width of

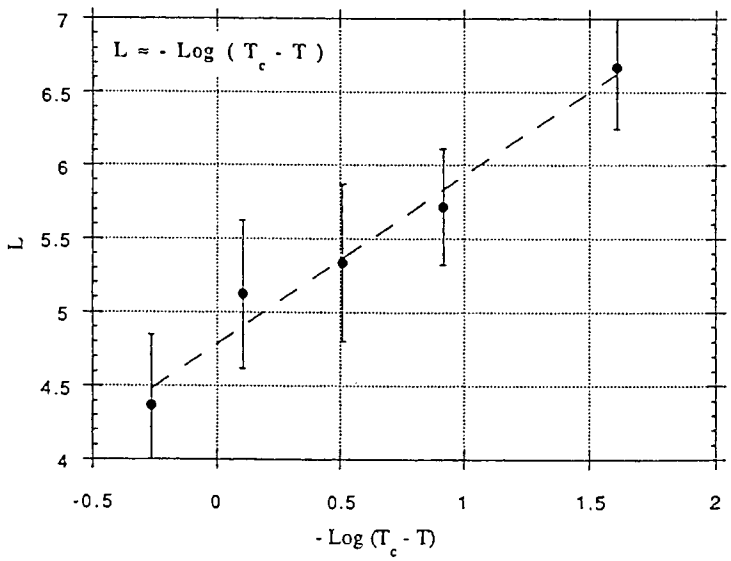


Figure 70. Variation of the width of the APBs in the vicinity of T_c, as a function of temperature. (Courtesy of C. Ricolleau [259].)

the domain walls [259]. Although the concentration of the alloy was chosen to avoid two-phase effects, the Cu–17% Pd system still presents a two-phase field about 1° wide. The presence of this two-phase field hampers the boundary-width measurements as well as the analysis of interface instabilities in the immediate vicinity of T_c; on the other hand, the measurements clearly allow us to distinguish between the wetting regime and the two-phase regime. In order to measure the exact width of the APB a number of precautions have to be taken. However, some problems persist: (a) the width of the interface is always measured at room temperature after quenching, (b) high-resolution images invariably show a sharper interface than conventional dark field images, and (c) close to T_c the APBs show a tendency to fluctuate dramatically and they may indicate the initial stages of formation of a long-period superstructure. The results of the change in boundary width as function of temperature are shown in Fig. 70. They show a good agreement with statistical mechanics calculations which predict that the width of the interface logarithmically diverges as $\log(T_c - T)$ [260].

When the palladium content of the alloy is increased in the Cu–Pd system, the $L1_2$ structure becomes unstable towards APB formation, and a one-dimensional long-period superstructure is formed. The average spacing between subsequent interfaces increases and is a function of composition and temperature [248, 261]. Close to the disordering temperature the interfaces exhibit similar wetting phenomena as the individual boundaries; this is clear from the image of Fig. 71, which is from a Cu–19 at.% Pd alloy annealed only 5°C below the disordering temperature. The disordered region is 1–2 $L1_2$ unit cells wide and, moreover, the interfaces, which at lower temperature are strictly bound to (001) planes, now become more or less wavy. The inset to Fig. 71, which is a processed

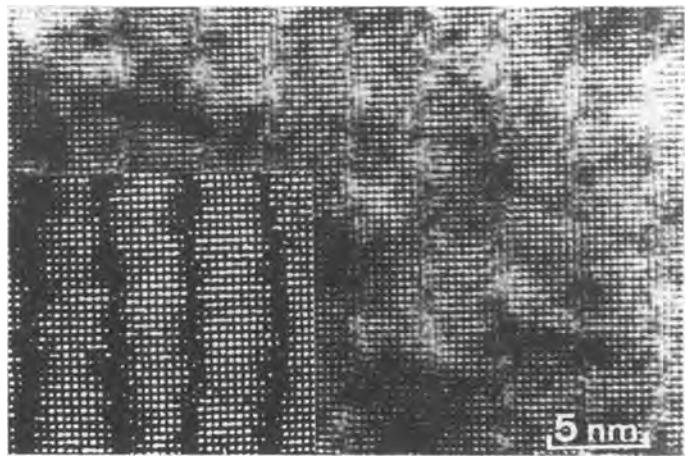


Figure 71. High-resolution image along [010] of the incommensurate long-period structures in $Cu_{81}Pd_{19}$; the processed image is shown as an inset.

image of the original HREM, accentuates the local changes at the interface. These effects are a general feature of long-period superstructures which are formed from a solid solution.

1.6.3 Displacive Transformations in Metals and Alloys

Microscopic investigations have played and are still playing a crucial role in the understanding of the different aspects of martensitic and other displacive transformations. These aspects cover a whole range of length-scales going from meso-scale surface configurations to atomic scale lattice structures. As a result, these studies have included the use of a large variety of microscopy techniques such as OM, SEM, TEM and HREM, each with its own aim and resolution. Also, recent advances mainly in the field of SP microscopy techniques have been applied with success, although the interpretation of images obtained by the latter is still subject to a lot of discussion.

The main points of interest in the study of martensitic transformations can be related to the crystallographic relation between parent and product phases, the microstructure of the product phase, the existence and atomic structure of internal defects and interfaces, the atomic and domain configurations leading to tweed or other precursing patterns and the understanding of the dynamics of the transformation. In the following, some typical results of microscopic investigations on these topics will be presented.

1.6.3.1 Microstructures

As a result of the strain-driven character and a lowering in symmetry during martensitic transformations, typical microstructures and well-defined habit planes and interfaces minimizing the strain energy are formed. The study of these configurations involves the observation of groups of martensite plates and the determination of their relative orientation relationships as well as that with the parent phase. The choice of microscopy technique

will depend on the length-scales of these microstructures. In the case of OM the orientation of the different structures or variants has to be determined separately by, for example, back-reflection X-ray Laue diffraction. In most cases the latter experiments involve the use of single-crystal parent phase material. Some recent examples of such an experiment are shown in Fig. 72 where an optical microscope is placed above a flat single crystal of Cu–14.0 wt.% Al–3.9 wt.% Ni with its parent phase surface normal along a (a) [111] and (b) a [100] axis [262]. The transformation in Fig. 72a is induced by in situ cooling

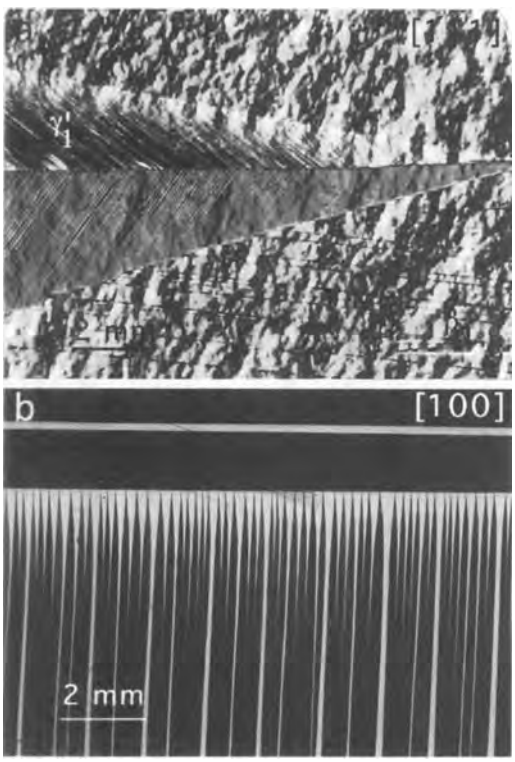


Figure 72. OM images of the β_1 to γ_1' martensitic transformation in Cu–14.0 wt.% Al–3.9 wt.% Ni. (a) Wedge-like microstructure with type II twins and (b) compound twins of different variants. (Courtesy of R. James [262].)

without stress, while in that in Fig. 72b a biaxial stress along [01$\bar{1}$] and [011], that is, normal to the edges of the specimen, was added. In Fig. 72a a wedge-like martensite microstructure with internal type II twinning and embedded in the austenite is observed, a configuration which can only be expected for special values of the lattice parameters [263]. By carefully monitoring the tension strengths, the sharp habit plane can be seen to proceed through the crystal. Figure 72b shows compound twins of which the volume ratio can be changed by cycling the specimen and changing the respective loads on the different stress axes. Also, twin layering and branching can be observed in the martensite. Due to accurate knowledge of the experimental conditions, the observations and measurements in this type of experiment can easily be compared with theoretical predictions, in this case of the newly developed non-linear elastic theory [264]. In order to obtain good contrast the surface has to be properly polished while the material is in the parent phase.

When the length-scale of the microstructures becomes smaller, for example, below 100 μm, a probe with shorter wavelength has to be used, and one is usually referred to the different modes of electron microscopy. The possibility of combining conventional imaging, transmission as well as scanning, with diffraction in a given experimental set-up is very powerful. Although many materials have been investigated in this way, new cases are still being discovered and analyzed. In Ni–Al, for example, Maxwell and Grala for the first time used the term 'martensite-like' in a technical report in 1954 [265] showing OM images indeed revealing the typical microstructures that were

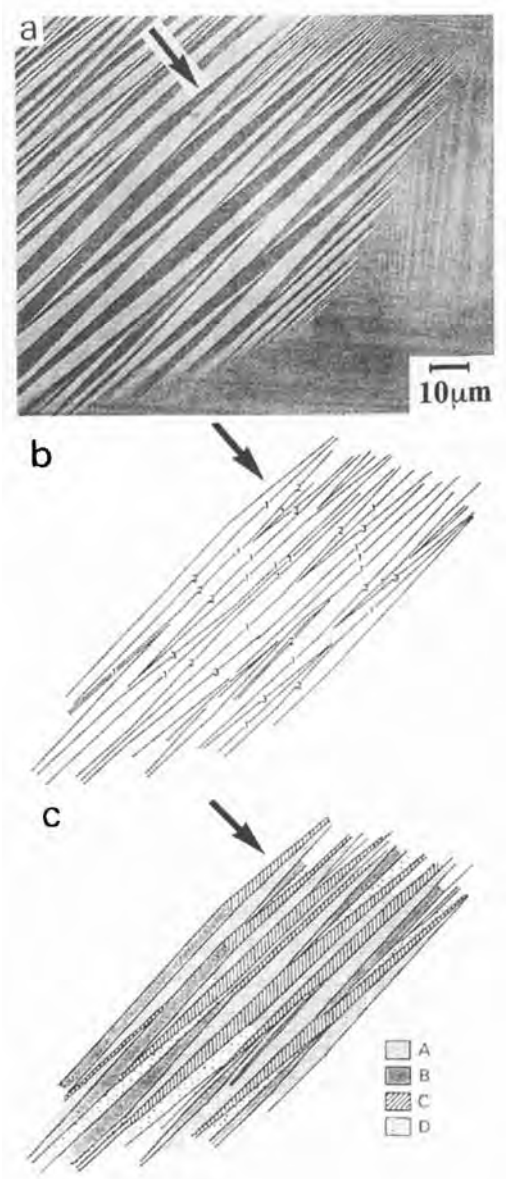


Figure 73. (a) SEM image of a self-accommodating group of 14M Ni–Al martensite. X-ray diffractometry yields the different orientations of the plates which reveal (b) three orientations of twin boundaries and (c) four habit plane variants of plates. (Courtesy of Y. Murakami et al. [267].)

characterized before as belonging to brass martensites. The first extensive report on the crystallography of the 2M or 3R martensite structure in Ni–Al was presented by Chakravorty and Wayman [266], using a combination of optical, interference and TEM techniques. Careful investigations of the traces of habit planes and plate interfaces yielded experimental values that could be compared with theoretical predictions based on strain energy minimization in linear elastic theory [227], confirming the influences of elastic strains on the microstructure of the martensite. Recently, Murakami et al. revisited this system and discussed the self-accommodation of martensite plates of the 14M structure, which is a long-period microtwinned version of the 2M structure (see also below) [267]. In Fig. 73 a SEM image and corresponding interpretation based on X-ray diffractometry of such a group is shown. Later, the same group also showed a perfect match between the theoretical predictions and experimental measurements on habit planes and the orientation relationship for the same martensite structure but stress induced [268].

1.6.3.2 Internal and Atomic Structure

When the interest lies in the internal or atomic structure of the martensite plates, one has to turn to higher magnifications including HREM studies. Again for binary Ni–Al, Mohanty and Rath presented the first transmission electron micrographs of the martensite phase in this system [269] at the 30th Annual EMSA meeting in 1972, their pictures clearly showing heavy faulting inside the

martensite plates which were identified as $\{111\}_{f.c.t.}$ twins. These twins are on close packed planes of the martensite $L1_0$-type structure, resulting from the tetragonal distortion of the B2 CsCl-type austenite. When the resolution of the instruments was improved, the regularity of these twins could be investigated in more detail. Schryvers combined data on the distribution of microtwins as gathered by image processing of TEM and HREM images from different martensite plates with different length-scales in order to obtain the surface energy of the microtwin planes after fitting with theoretical predictions [270]. A similar study in Ni–Mn by Baele et al. showed a correlation between the composition, namely the e/a ratio, and the microtwin periodicity [271]. Using selected area electron diffraction (SAED) in combination with one-dimensional line resolution, Chandrasekaran et al. observed several microtwin stacking sequences in local regions of Ni–Al martensite [272, 273]. The most frequently observed regular sequence in this system, however, has a 14-layered sequence indicated as 14M (or 7R in earlier publications), which was first detected by Reynaud in an in situ SAED cooling experiment [274]. Later, the actual stacking sequence of $(5\bar{2})_2$, first suggested by Martynov et al. on the basis of X-ray diffraction [275], was directly confirmed by Schryvers et al. using HREM [276], and is seen in Fig. 74. Recently, new stacking variants such as 14M, 12M and 10M have been reported in ternary alloys of Ni–Al–Mn by Inoue et al. [277].

A complementary technique to HREM imaging is the analytical site determination technique ALCHEMI in which the site occupation for a given atom type can

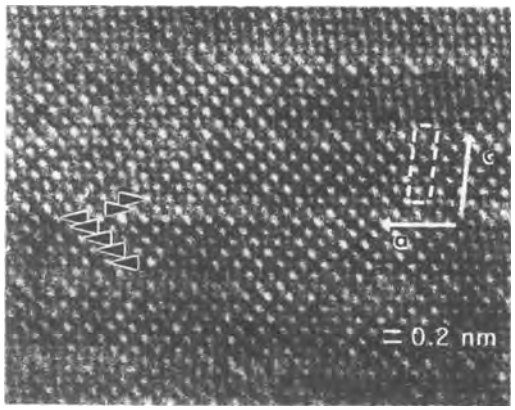


Figure 74. HREM image showing the $(5\bar{2})$ stacking of the 14M structure in $Ni_{62.5}Al_{37.5}$ [276].

be determined by taking energy dispersive X-ray spectroscopy (EDX) patterns under different excitation conditions. An example is shown in Fig. 75 where positive and negative excitation errors with respect to the 200 reflection give a different EDX

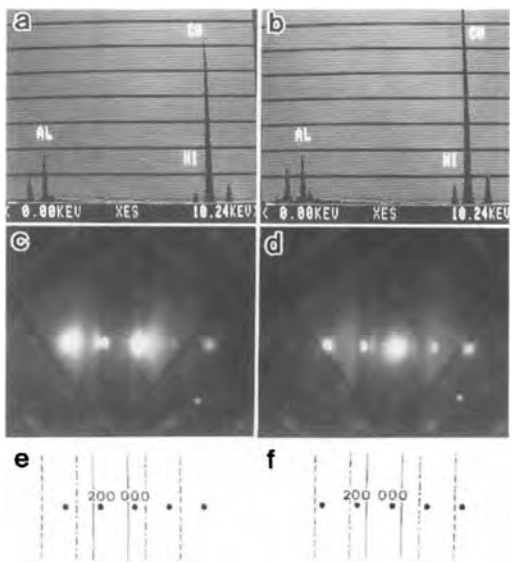


Figure 75. EDX spectra taken under different excitation conditions ((a) $s > 0$ and (b) $s < 0$) in Cu–Al–Ni from which a nickel occupation of the pure copper sublattice was concluded. (Courtesy of Y. Nakata et al. [278].)

signal for the nickel atoms in a Cu–Al–Ni parent DO_3-type phase. From this and related spectra from differently treated samples it was concluded that the nickel atoms occupy the pure copper plane in this structure and that the rise of M_s with aging of the parent state is attributed to the progress of the second nearest-neighbor ordering between aluminum and copper atoms [278].

1.6.3.3 Interfaces, Defects and Precipitates

Using HREM, irregularities in the stacking as well as local defects such as ledges on a twin interface or stacking fault can be investigated. Such phenomena often occur close to large-scale defects such as grain boundaries, precipitates or even the thinned edge of a TEM sample. In the latter case, the specific orientation of the surface might induce thin foil effects which should be taken into account when interpreting the micrographs. An example of this is shown in Fig. 76, where the microtwin width increases when approaching the edge of the foil [276]. In the above examples the atomic structure was always considered to be ideal irrespective of whether one looks far away from or close to a twin or habit plane. An extensive HREM study on nickel–manganese indeed could not prove any distortion near these boundaries.

Crystallographic information of single defects can also be obtained by large angle convergent beam electron diffraction (LACBED). An example is the determination of the Burgers vector $\boldsymbol{b}$ of dislocations in the austenite of a Ni–Ti–Co alloy as shown in Fig. 77 [279–281]. Here the Fourier and direct space information is

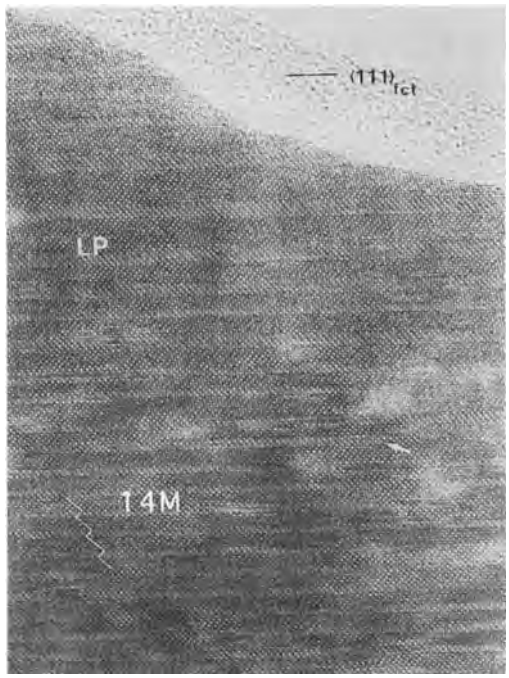


Figure 76. Effect of foil thickness on the microtwin width in Ni–Al martensite [276].

combined in a single micrograph (obtained at the objective back focal plane) where the cross-overs of dark Bragg lines with an oriented dislocation line produces fringes at the intersections. The number of fringes n follows the formula $n = \boldsymbol{g} \cdot \boldsymbol{b}$, with $\boldsymbol{g}$ the

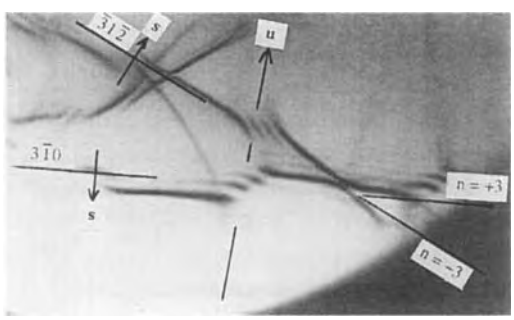


Figure 77. Burgers vector determination by LACBED in Ni–Ti–Co ($\boldsymbol{u}$, direction dislocation line; $\boldsymbol{s}$, indicates sign of excitation error outside Bragg line; $\boldsymbol{b}$, [100]). (Courtesy of J. Pons et al. [279].)

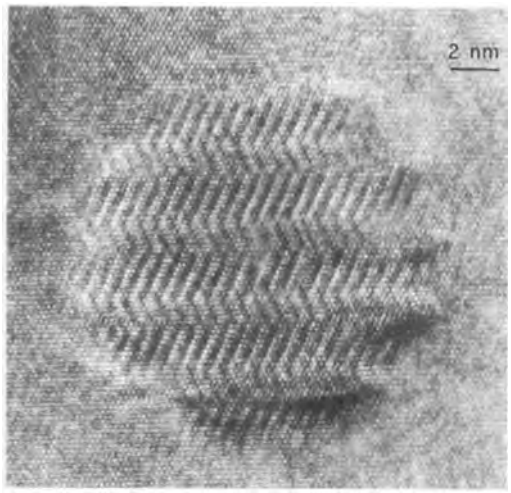


Figure 78. HREM image of a twinned 9R copper precipitate in a Fe–1.3 wt.% Cu alloy aged at 550°C for 30 h. (Courtesy of P. Othen et al. [282].)

Bragg line indices, from which the Burgers vector can be determined when using at least two independent Bragg lines.

In some cases the displacive transformation only occurs in precipitates distributed in a matrix. In alloys, an extensively studied example is that of copper precipitates in thermally aged or electron irradiated Fe–Cu material in which these precipitates undergo a transformation, thought to be martensitic, from b.c.c. to a 9R twin stacking as shown in Fig. 78 [282, 283]. It was found that the transformation occurs for a particle size below 4–5 nm, after which they grow by adding twin planes. A very important advantage of these type of electron microscopic observations is that one immediately recognizes the orientation relationship between the parent and product phases as well as that with the surrounding matrix. Moreover, the well-defined and observable constraints on the interface can yield information difficult to obtain in ordinary martensitic transformations.

1.6.3.4 Precursor Phenomena

In many of the alloy systems exhibiting a martensitic transformation a characteristic so-called tweed pattern is observed in two-beam TEM images of the austenite as shown in Fig. 79, including the corresponding SAED pattern showing the diffuse streaks at the Bragg reflections. These patterns can exist up to several decades of degrees above the transition temperature M_S. Such patterns were first explained in detail by Tanner for the case of GP zones in Cu–Be [284]. An extensive study on the imaging characteristics of this tweed pattern in Ni–Al was conducted by Robertson and Wayman [285]. As this contrast appears under two-beam conditions, the actual choice of diffracting vector as well as the value of the extinction distance s play an important role in the

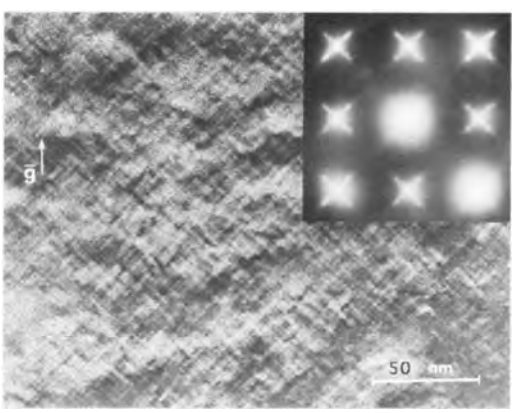


Figure 79. Conventional two-beam tweed pattern in Ni–Al showing ⟨110⟩ striations as a result of soft transverse modes in the matrix responding to local strain centers. The inset shows the corresponding diffuse intensity around the Bragg reflections.

final contrast, for example, affecting its orientation and coarseness. The origin of the image contrast is believed to lie in the existence of strain modulations of the matrix, appearing in more than one but usually equivalent crystallographic directions. Since then, tweed patterns have been found in many different systems including high-T_c superconductors and other ceramics [286]. The actual structural origin of the strain modulation will depend on the system under consideration. Valid candidates are point defect strain centers in rapidly quenched alloys, preliminary clustering in ordering systems, GP zones when decomposition occurs, and so forth.

In the case of Ni–Al, HREM provided some more detailed information on the atomic configurations underlying the tweed contrast. Two distinct but typical examples of such images are shown in Fig. 80 [287], obtained from thin but flat regions of a wedge-shaped foil. In Fig. 80a the local distortions are clearly recognized as patches in which the expected two-dimensional square white dot pattern corresponding with the undistorted ordered structure [238, 239] is changed into line patterns along two possible orientations. Optical diffractograms (ODs) of large areas from such HREM micrographs immediately reveal the static or dynamic nature of the tweed pattern. If the OD reveals the same diffuse intensity streaks as the SAED patterns, as seen in Fig. 80c, the corresponding distortions are recorded in the image (illumination time around 1 s), indicating static displacements [288]. In Fig. 80b, contrast modulations parallel with {110} b.c.c. planes and with a wavelength of ±1.3 nm are superposed onto the white dot pattern. In some domains a homogeneous shear could also be

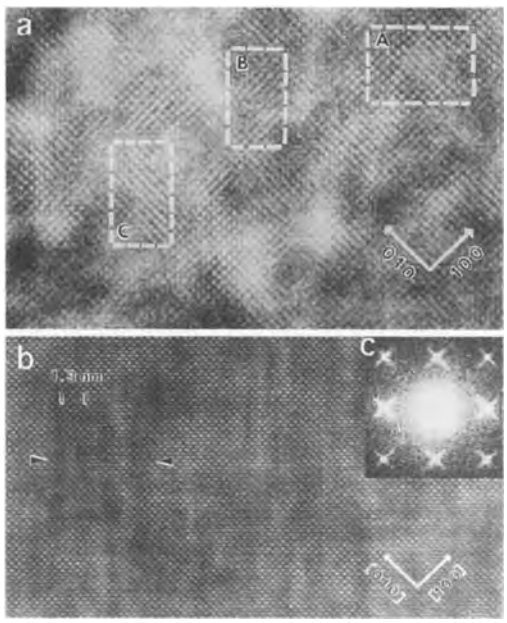


Figure 80. HREM image revealing domains of micromodulations observed above M_s in Ni–Al: (a) only homogeneous distortions of different variants are visible, (b) a 1.3 nm contrast modulation along ⟨110⟩ is superposed [287], and (c) OD from a HREM image of (b) showing diffuse intensity at the Bragg reflections indicating the static nature of the distortions [288].

measured. Each of these *micro*modulations has a correlation length of a few nanometers, as indicated by the arrows on one domain. Based on image simulations, these micromodulations were interpreted as due to transverse sine waves superposed onto a homogeneous shear [287]. Since in this multibeam imaging condition one is looking along a well-chosen zone axis, the observed micromodulations and shears are intrinsic to the material and not a result of the imaging conditions, as is the case in the conventional tweed images of Fig. 79. The phenomenon yielding the latter is thought to be correlated to the strains between the different domains. To what extent these

precursing micromodulations and shears influence or are necessary for the ensuing transformation is still a matter of debate [289].

Another example of a new insight into precursors as obtained by HREM is found in rapidly quenched Ti–8 at.% Mo where the primary direction of alignment in the precursors strongly deviates from the final ⟨111⟩ direction necessary to form the ω structure [290]. Although a valid suggestion for the atomic configuration underlying the HREM images could be presented, a proper explanation for the appearance of this deviation is still at large.

1.6.3.5 Nucleation and Transformation Characteristics

The above examples of the use of different microscopy techniques in the study of displacive transformations mainly refer to problems related to static situations. In order to understand the nucleation mechanisms and the actual behavior of the lattice during the transformation, however, in situ experiments have to be performed. In the case of structural transformations this implies the use of heating and cooling as well as straining stages or even, as recently developed for electron microscopy, a combination of both [291]. Unfortunately, working in situ still decreases the obtainable resolution of the entire set-up so that only conventional TEM and SAED are possible, except for some very specific cases.

A typical example of a conventional in situ TEM study is shown in the sequence of Fig. 81 where the nucleation of the R phase in cycled $Ti_{50}Ni_{48}Al_2$ material is followed [292]. It is seen that the dislocations in the B2 parent phase act as nucleation centers for the product phase, of which the structure was determined using SAED (see inset). It is interesting to see that a single R plate can form after the nucleation of different plates A and B

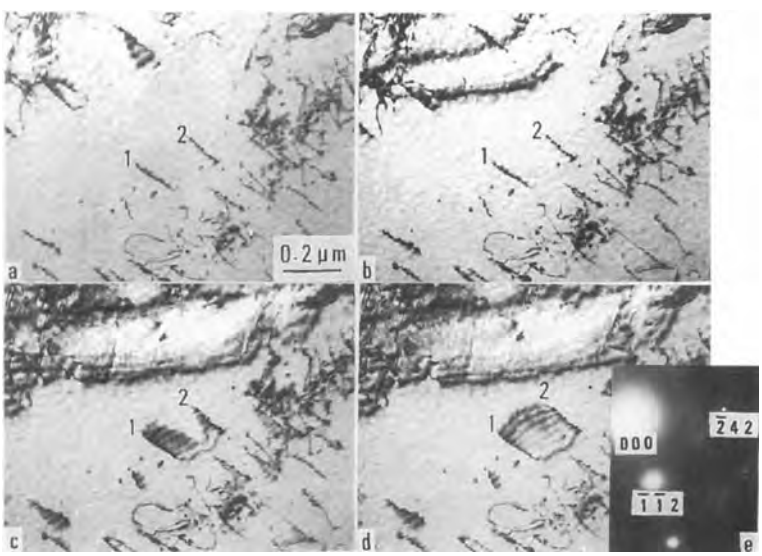


Figure 81. In situ cooling sequence showing the nucleation and growth of R phase plates in $Ti_{50}Ni_{48}Al_2$. Note the formation of a single plate between two separate dislocations. (Courtesy of T. Saburi [292].)

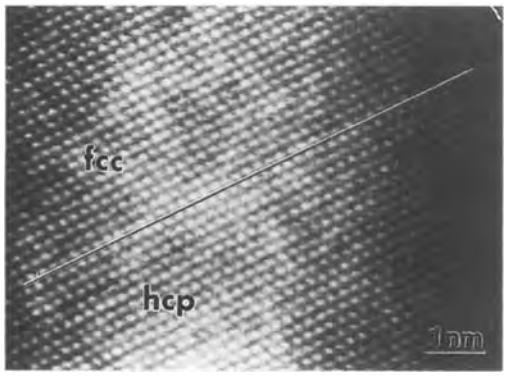


Figure 82. In situ HREM image of a coherent and atomically flat $\{111\}_{f.c.c.}//(0001)_{h.c.p.}$ interface observed in a $\langle 110\rangle_{f.c.c.}//\langle 11\bar{2}0\rangle_{h.c.p.}$ orientation in Co–32 wt.% Ni. (Courtesy of J. Howe [293, 294].)

(Fig. 81c) at separate dislocations 1 and 2 (Fig. 81a). In the present experiment the transformation was steered by reducing (cooling) or increasing (heating) the beam intensity, and the appearance and disappearance of the R phase plates proceed in the same manner on repeated cooling and heating, as long as the cycling does not influence the dislocations.

As mentioned above, the use of in situ techniques usually limits the attainable resolution of the experiment so that in most studies only conventional microscopy can be used. However, a few examples from dedicated HREM instruments equipped with special in situ holders do exist today. In Fig. 82 a $\{111\}_{f.c.c.}//(0001)_{h.c.p.}$ interface is shown observed in a single crystal of Co–32 wt.% Ni and taken during an in situ heating experiment at about 100°C in a JEOL 4000EX equipped with a UHP40X hot-stage pole piece [293, 294]. In the present case the martensitic f.c.c. to h.c.p. (hexagonal close packing) transformation has a strong hysteresis around room temperature, enabling one to trap the system halfway through the transition and to study the dynamics of the transformation in the in situ stage. As seen in the image the interface is perfectly coherent and atomically flat.

In some cases, a martensite transition can be trapped inside the hysteresis loop, or a temperature or stress gradient is observed along the edge of a thinned electron microscopic sample. An example of the latter is shown in Fig. 83 which was found at a $\langle 110\rangle$ crack automatically formed during the last stages of electropolishing to an extremely thin foil [276]. SAED (see insets) shows that the wings around the crack are martensite (14M with and 2M without microtwinning) while the rest of the matrix is still austenite. However, when zooming in at the austenite close to the crack a gradual change in distortion is observed in HREM as well as SAED, attributed to the stress gradient [276]. Thus although these images were again obtained from a static situation, the actual information can immediately yield atomic detail on the transformation process.

1.6.4 Some Other Transformations in Alloys

The mere fact that the remaining transformations include diffusion as well as displacement of the atoms makes them interesting study cases for which a combination of different microscopy techniques is appropriate. In a detailed HREM plus electron microdiffraction study, Muto et al. showed that in the formation of the metastable Ni_2Al structure local reordering as well as atomic shuffles are involved [295, 296]. The latter

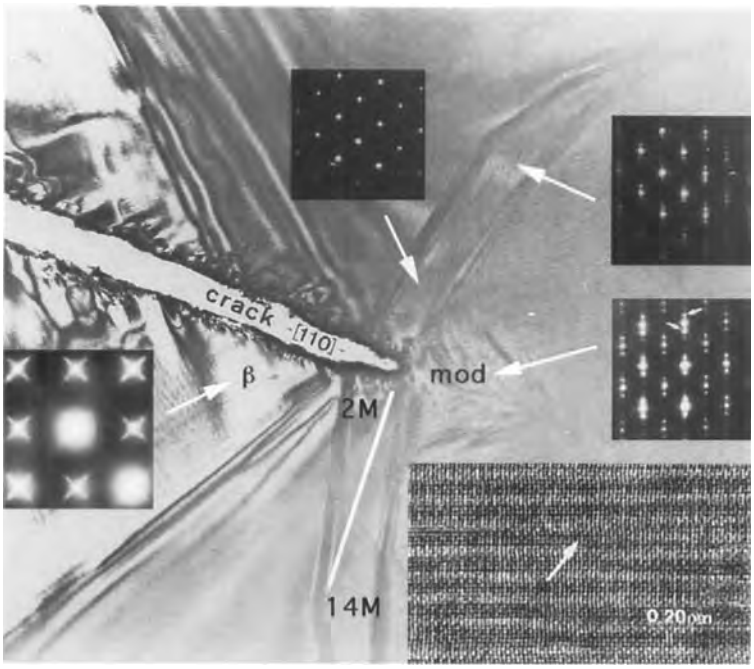


Figure 83. TEM image of a transformed matrix in the stress region surrounding a thin foil crack in $Ni_{62.5}Al_{37.5}$ [276]. HREM and SAED (insets) reveal 2M and 14M martensite in distinct regions as well as a broad area of decreasing strains and changing modulations between the martensite and austenite.

can be compared with ω-type shuffles, but appear to be only partial, that is, the collapse of $\{111\}_{B2}$ planes is incomplete ($z' < 0.083$). In Fig. 84 a comparison between experimental and calculated HREM images is shown, indicating that the shuffles cannot be larger than 50% of complete ω shuffles. On the other hand, the relative intensities of the super-reflections as measured by densitometry from electron microdiffraction patterns from single variants showed that these shuffles were at least 20% of complete ω shuffles. This is shown in Fig. 85. So although no exact atomic positions were obtained, the existence of a partial collapse was clearly established.

Another example of a combined transition is the precipitation of Ni_5Al_3 (5 : 3) by annealing the B2 Ni–Al matrix. One specific advantage of the usually slower transformation process of these coupled

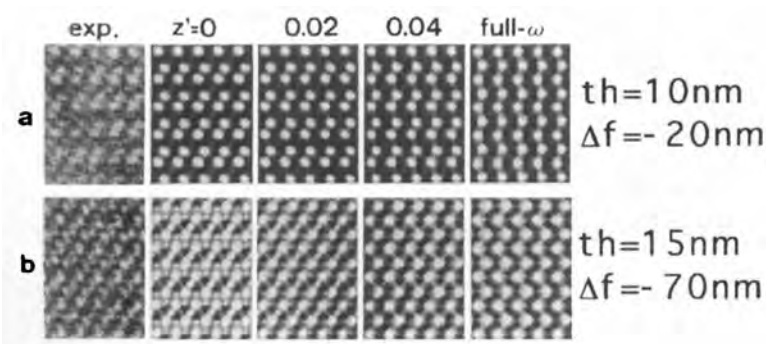


Figure 84. Experimental and simulated HREM of Ni_2Al showing the existence of a minimum amount of ω collapse [295].

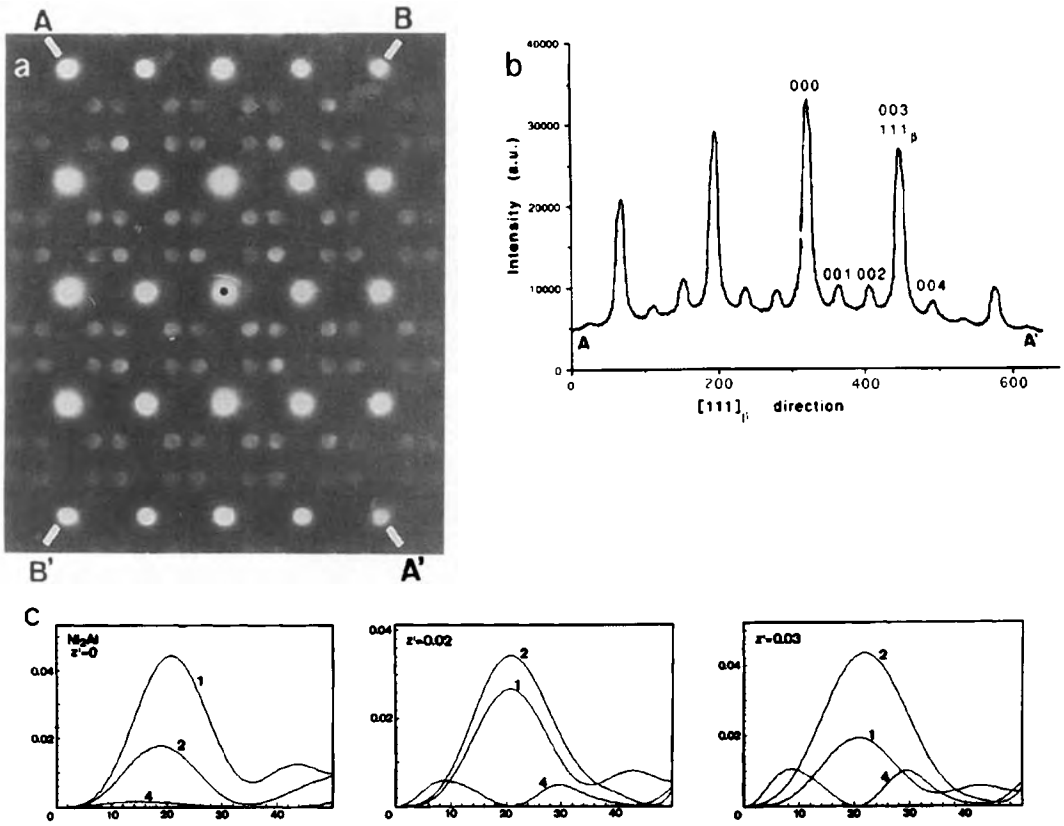


Figure 85. Comparison of (a) the microdiffraction pattern of a Ni_2Al precipitate and (b) its densitometry measurement with (c) corresponding simulations showing the incomplete nature of the ω collapse [295].

transformations when compared with pure displacive ones is the fact that intermediate transition stages are more readily observed. Due to the combination of the reordering process and the cubic to tetragonal distortion nanoscale star-shaped precipitates are formed as shown in Fig. 86 [297]. The internal atomic structure, the relative composition with respect to the matrix and the defect configuration of the interfaces can directly be investigated by HREM and nanoprobe EDX. Further on during the transformation these precipitates grow into twinned plates embedded in the matrix. From HREM images of the interface of such a plate with the matrix the orientation relationship as well as the lattice strain can be measured [298].

A classic case of coupled transformations is the class of Bainitic transformations in which martensite-like characteristics are combined with atom diffusion. Again, these transformations are usually much slower than pure martensitic ones, and intermediate situations are more easily retained in quenched samples. Moreover, in case the diffusion yields a decomposition, small-probe analytical microscopy will be of help to determine

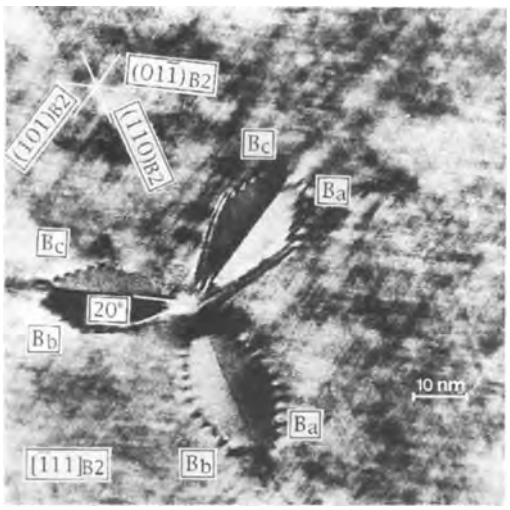


Figure 86. Ni_5Al_3 precipitate in a ⟨111⟩ B2 matrix formed by a twinning configuration of different variants indicated as B_a, B_b and B_c [297]. Due to severe distortions, dislocations still exist in the precipitate–matrix interface.

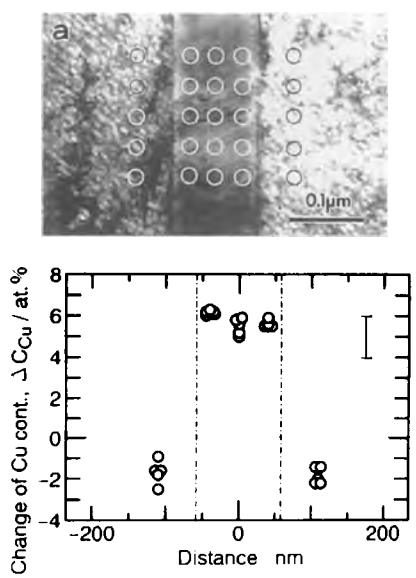


Figure 87. Relative copper content in a 100 nm thick Bainite plate in Cu–Zn–Au with respect to the surrounding matrix and obtained by EDX with a 20 nm diameter FEG probe. (Courtesy of T. Tadaki et al. [299].)

the nominal compositions of the Bainite and the surrounding matrix. An example is seen in Fig. 87, where a 100 nm thick Bainite plate in a Cu–Zn–Au alloy was analyzed with a 20 nm diameter probe from a field emission gun [299, 300]. The graph shows the relative increase of copper content in the plate with respect to the surrounding matrix and no differences between the central and edge regions of the plate. Although scanning TEM could provide a smaller probe, TEM was used since in this mode the exact location of the plate could best be observed.

1.6.5 Imaging Phase Transformations: Conclusions

In the present section we have selected some of the key contributions of microscopy techniques to the study of phase transformations in metals and alloys. Although it is of course impossible to give a complete survey of these in this extremely wide field of research, we have tried to cover most of the current topics of interest by showing examples from different materials and a variety of experimental techniques. Investigations on pretreated samples as well as in situ work are presented, and although the majority of cases were conducted using electron microscopy, from which different modes can yield information on different scales and of different type, some relevant results from other microscopy techniques have been presented.

With the further evolution of attainable resolution, the possibility to perform combined experiments and the improvements with new probe techniques, it is clear that

microscopy investigations on phase transformations in metals and alloys still have a bright future.

1.7 References

[1] P. B. Hirsch, A. Howie, R. B. Nicholson, D. W. Pashley, M. J. Whelan, *Electron Microscopy of Thin Crystals*, Butterworths, London **1965,** p. 156.

[2] L. Reimer, *Transmission Electron Microscopy*, Springer-Verlag, Berlin **1984**.

[3] G. Thomas, M. J. Goringe, *Transmission Electron Microscopy of Materials*, Wiley, New York **1979**.

[4] R. D. Heidenreich, *Fundamentals of Transmission Electron Microscopy of Materials*, Wiley, New York **1967**.

[5] A. P. Sutton, V. Vitek, *Phil. Trans. R. Soc. London A* **1983,** *309*, 1.

[6] J. Th. M. De Hosson, V. Vitek, *Philos. Mag. A* **1990,** *61*, 305.

[7] D. J. Holt, *J. Appl. Phys.* **1970,** *41*, 3197.

[8] E. C. Aifantis, *Mater. Sci. Eng.* **1986,** *81*, 563.

[9] J. J. Kratochvil, *Scripta Metall.* **1990,** *24*, 891.

[10] C. Kittel, *Introduction to Solid State Physics*, 6th edn, Wiley, New York **1986**.

[11] D. J. H. Cockayne, in *Diffraction and Imaging Techniques in Materials Science* (Ed.: S. Amelinckx), North-Holland, Amsterdam **1978,** p. 153.

[12] M. H. Loretto, R. E. Smallman, *Defect Analysis in Electron Microscopy*, Chapman and Hall, London **1975**.

[13] J. P. Hirth, J. Lothe, *Theory of Dislocations*, McGraw-Hill, New York **1968**.

[14] J. W. Steeds, *Anisotropic Elasticity Theory of Dislocations*, Clarendon Press, Oxford **1973**.

[15] J. Van Landuyt, R. Gevers, S. Amelinckx, *Phys. Stat. Sol.* **1965,** *10*, 319.

[16] J. W. Edington, *Practical Electron Microscopy in Materials Science*, Macmillan Press, London **1975**.

[17] M. Rühle, M. Wilkens, *Proc. 5th European Congr. On Electron Microscopy* **1972,** p. 416.

[18] J. A. Chandler, *Practical Methods in Electron Microscopy 5-II* (Ed.: A. M. Glauert), North-Holland, Amsterdam **1973/1977**.

[19] P. M. Kelly, J. Nutting, *Laboratory Workers Handbook* **1964,** 4.

[20] G. Castaing, *Rev. Met.* **1955,** *52*, 669.

[21] H. De Beurs, J. Th. M. De Hosson, *Appl. Phys. Lett.* **1988,** *53*, 663.

[22] H. De Beurs, J. A. Hovius, J. Th. M. De Hosson, *Acta Metall. Mater.* **1988,** *36*, 3123.

[23] J. Noordhuis, J. Th. M. De Hosson, *Acta Metall. Mater.* **1992,** *40*, 3317.

[24] J. Th. M. De Hosson, J. Noordhuis, B. A. van Brussel, *Mater. Sci. Forum* **1992,** *102–104*, 393.

[25] F. Z. Cui, A. M. Vredenberg, F. W. Saris, *Appl. Phys. Lett.* **1988,** *53*, 2152.

[26] H. G. Feller, R. Klinger, W. Benecke, *Mater. Sci. Eng.* **1985,** *69*, 173.

[27] H. J. Hegge, H. De Beurs, J. Noordhuis, J. Th. M. De Hosson, *Met. Trans.* **1990,** *21A*, 987.

[28] B. A. Van Brussel, J. Noordhuis, J. Th. M. De Hosson, *Scripta Metall. Mater.* **1991,** *25*, 1719.

[29] R. W. Balluffi, A. V. Granato, in *Dislocations in Solids* (Ed.: F. R. N. Nabarro), North-Holland, Amsterdam **1979,** p. 1 (volume 4).

[30] J. S. Koehler, *Phys. Rev.* **1969,** *181*(3), 1015.

[31] H. De Beurs, J. Th. M. De Hosson, *Scripta Metall. Mater.* **1987,** *21*, 627.

[32] J. Noordhuis, J. Th. M. De Hosson, *Acta Metall. Mater.* **1990,** *38*, 2067.

[33] U. F. Kocks, *Mater. Sci. Eng.* **1977,** *27*, 291.

[34] J. W. Martin, *Micromechanisms in Particle-Hardened Alloys*, Cambridge University Press, Cambridge **1980**.

[35] V. Gerhold, H. Haberkorn, *Phys. Status Solidi* **1966,** *16*, 675.

[36] A. J. Ardell, *Met. Trans.* **1985,** *16A*, 2131.

[37] K. Hanson, J. W. Morris, *J. Appl. Phys.* **1975,** *46*, 2378.

[38] N. E. W. Hartley, *J. Vac. Sci. Technol.* **1975,** *12*, 485.

[39] E. Gerritsen, J. Th. M. De Hosson, *Fundamental Aspects of Inert Gases in Solids* (Eds.: S. E. Donnely, J. E. Evans), Plenum Press, New York **1991**, p. 153.

[40] W. G. Wolfer, *Phil. Mag.* **1989,** *A59*, 87.

[41] J. H. Evans, *J. Nucl. Mater.* **1977,** *68*, 129.

[42] W. G. Wolfer, *Sandia Report 87-8859, Sandia Natl. Lab.*, Sept. **1987**.

[43] E. Johnson, U. Littmark, A. Johansen, C. Christodoulides, *Phil. Mag.* **1982,** *A45*, 803.

[44] E. Johnson, A. Johansen, L. Sarholt-Kristensen, L. Gråbæk, N. Hayashi, I. Sakamoto, *Nucl. Instrum. Methods* **1987,** *B19/20*, 171.

[45] S. Kajiwara, *Met. Trans.* **1986,** *17A*, 1693.

[46] J. H. Evans, *Nucl. Instrum. Methods* **1986,** *B18*, 16.

[47] P. J. Goodhew, *Scripta Metall. Mater.* **1984,** *18*, 1069.

[48] P. E. Cary, in *ICSP-1* (Ed.: A. Niku-Lari), Pergamon Press, Oxford **1981,** p. 23.

[49] D. W. Hammond, S. A. Meguid in *Surface Engineering* (Ed.: S. A. Meguid), Elsevier Applied Science, London **1990,** p. 386.
[50] B. P. Bardes (Ed.) *Metals Handbook*, 9th edn, American Society for Metals, Ohio **1982,** Vol. 5.
[51] H. Jiawen in *ICSP-2* (Ed.: H. O. Fuchs), Paramus, New Jersey **1984,** p. 266.
[52] R. O. Ritchie, *Fundamentals of Tribology*, MIT Press, Cambridge, MA **1978**, p. 127.
[53] B. Syren, H. Wohlfahrt, E. Macherauch, *Proc. 2nd Int. Conf. on Mechanical Behaviour of Materials, Boston* **1976,** p. 212.
[54] W. Luo, B. Noble, R. B. Waterhouse in *Impact Surface Treatment* (Ed.: S. A. Meguid), Elsevier Applied Science, London **1986,** p. 57.
[55] L. Clapham, R. W. Smith, *Acta Metall. Mater.* **1989,** *37*, 303.
[56] H. J. Hegge, J. Th. M. De Hosson, *Acta Metall. Mater.* **1990,** *38*, 2471.
[57] J. L. Murray, A. J. McAlister, *Bull. Alloy Phase Diagrams* **1984,** *5*, 75.
[58] G. J. Den Otter, A. van den Beukel, *Phys. Status Solidi A* **1979,** *55*, 785.
[59] F. R. N. Nabarro, *J. Less Common Met.* **1972,** *28*, 257.
[60] J. Th. M. De Hosson, G. Boom, U. Schlagowski, O. Kanert, *Acta Metall. Mater.* **1986,** *34*, 1571.
[61] R. L. Fleischer, *Acta Metall.* **1961,** *9*, 996.
[62] R. L. Fleischer, *Acta Metall.* **1963,** *11*, 203.
[63] H. S. Chen, A. Acrivos, *Int. J. Solids Structures* **1978,** *14*, 349.
[64] S. R. MacEwen, O. A. Kupcis, B. Ramaswami, *Scripta Metall. Mater.* **1969,** *3*, 441.
[65] E. Orowan, *Symp. on Internal Stresses in Metals and Alloys*, Institute of Metals, London **1948,** p. 451.
[66] J. W. Christian, P. R. Swann, *Met. Soc. Conf.* **1963,** *29*, 105.
[67] G. K. Williamson, R. E. Smallman, *Phil. Mag.* **1956,** *1*, 34.
[68] R. E. Stolz, R. M. Pelloux, *Met. Trans.* **1976,** *7A*, 1295.
[69] L. F. Mondolfo, *Aluminum Alloys. Structure and Properties* **1976,** Butterworth, London. p. 264.
[70] G. W. Lorimer in *Precipitation Processes in Solids* (Eds.: K. C. Russell, H. I. Aaronson), AIME, New York **1978,** p. 87.
[71] M. F. Ashby, K. E. Easterling, *Acta Metall. Mater.* **1984,** *32*, 1935.
[72] S. Amelinckx, W. Bontinck, W. Dekeyser, F. Seitz, *Phil. Mag.* **1957,** *2*, 355.
[73] P. M. Kelly, *Scripta Metall. Mater.* **1972,** *6*, 647.
[74] P. Merle, F. Fouquet, J. Merlin, *J. Mater. Sci. Eng.* **1981,** *50*, 215.
[75] J. D. Embury in *Strengthening Methods in Crystals* (Eds.: A. Kelly, R. B. Nicholson), Wiley, New York **1971,** p. 331.
[76] M. G. Scott, J. A. Leake, *Acta Metall. Mater.* **1975,** *23*, 503.
[77] K. Aoki, O. Izumi, *Nippon Kinzaku Gakkaishi* **1979,** *43*, 1190.
[78] I. Baker, E. M. Schulson, J. A. Horton, *Acta Metall. Mater.* **1987,** *35*, 1533.
[79] S. Hanada, T. Ogura, S. Watanabe, O. Izumi, T. Masumoto, *Acta Metall. Mater.* **1986,** *34*, 13.
[80] J. J. Hauser, B. Chalmers, *Acta Metall. Mater.* **1961,** *9*, 802.
[81] W. D. Brentnall, W. Rostoker, *Acta Metall. Mater.* **1965,** *13*, 187.
[82] D. J. Dingley, R. C. Pond, *Acta Metall. Mater.* **1979,** *27*, 667.
[83] M. Elkajbaji, J. Thibault-Desseaux, *Phil. Mag. A* **1988,** *58*, 325.
[84] M. L. Kronberg, F. H. Wilson, *Trans. Metall. Soc. AIME* **1949,** *185*, 501.
[85] R. W. Balluffi, A. Brokman, A. H. King, *Acta Metall. Mater.* **1982,** *30*, 1453.
[86] H. J. Frost, *Acta Metall. Mater.* **1987,** *35*, 519.
[87] W. Bollmann, *Crystal Defects and Crystalline Interfaces*, Springer-Verlag, Berlin **1970**.
[88] A. H. King, D. A. Smith, *Acta Crystallogr. A* **1980,** *36*, 335.
[89] C. T. Forwood, L. M. Clarebrough, *Electron Microscopy of Interfaces in Metals and Alloys*, Adam Hilger, Bristol **1991**.
[90] D. H. Warrington, *Grain-Boundary Structure and Kinetics*, American Society for Metals, Metals Park, Ohio, USA **1980**.
[91] W. Bollman, *Crystal Lattices, Interfaces, Matrices* **1982**.
[92] T. Schober, R. W. Balluffi, *Phil. Mag.* **1971,** *24*, 165.
[93] E. S. P. Das, E. S. Dwarakadasa, *J. Appl. Phys.* **1974,** *45*, 574.
[94] W. Bollmann, B. Michaut, G. Sainfort, *Phys. Status. Solidi. A* **1972,** *13*, 637.
[95] P. A. Doyle, P. S. Turner, *Acta Crystallogr. A* **1968,** *24*, 390.
[96] C. A. P. Horton, J. M. Silcock, G. R. Kegg, *Phys. Status Solidi A* **1974,** *26*, 215.
[97] R. C. Pond, D. A. Smith, V. Vitek, *Scripta Metall.* **1978,** *12*, 699.
[98] J. S. Liu, R. W. Balluffi, *Phil. Mag. A* **1985,** *52*, 713.
[99] L. M. Clarebrough, C. T. Forwood, *Phys. Status Solidi A* **1980,** *58*, 597.
[100] Z. Shen, R. H. Wagoner, W. A. T. Clark, *Scripta Metall.* **1988,** *20*, 921.
[101] Z. Shen, R. H. Wagoner, W. A. T. Clark, *Acta Metall.* **1986,** *36*, 3231.

[102] T. C. Lee, I. M. Robertson, H. K. Birnbaum, *Scripta Metall.* **1989,** *23*, 799.
[103] L. P. Kubin, P. Veyssière, *Proc. 10th Int. Congr. on Electron Microscopy, Hamburg*, Deutsche Gesellschaft für Elektronenmikroskopie e.V. (Battelle-Institut e.V.), Frankfurt/Main **1982,** p. 531.
[104] G. L. Bleris, J. G. Antonopoulos, Th. Karakostas, P. Delavignette, *Phys. Status Solidi A* **1981,** *67*, 249.
[105] G. M. Bond, I. M. Robertson, H. K. Birnbaum, *J. Mater. Res.* **1987,** *2*, 436.
[106] H. R. Pak, T. Saburi, S. Nenno, *Scripta Metall. Mater.* **1976,** *10*, 1081.
[107] F. D. Tichelaar, P. H. H. Rongen, Y. G. Zhang, F. W. Schapink, *Interface Sci.* **1994,** *2*, 105.
[108] P. Veyssiere, J. Douin, P. Beauchamp, *Phil. Mag. A* **1985,** *51*, 469.
[109] I. Baker, E. M. Schulson, *Phys. Status Solidi A* **1984,** *85*, 481.
[110] B. H. Kear, J. M. Oblak, A. F. Giamei, *Metall. Trans.* **1970,** *1*, 2477.
[111] W. A. T. Clark, R. H. Wagoner, *Dislocations in Solids* (Eds.: H. Suzuki, T. Ninomiya, K. Sumino, S. Takeuchi), University of Tokyo Press, Tokyo **1985,** p. 647.
[112] T. C. Lee, I. M. Robertson, H. K. Birnbaum, *Metall. Trans. A* **1990,** *21*, 2437.
[113] F. D. Tichelaar, *Thesis, University of Delft* **1992**; F. D. Tichelaar, F. W. Schapink, *Phil. Mag. A* **1990,** *62*, 53.
[114] B. P. Pestman, J. Th. M. De Hosson, V. Vitek, F. W. Schapink, *Phil. Mag. A* **1991,** *64*, 951.
[115] J. A. Horton, M. K. Miller, *Acta Metall. Mater.* **1987,** *35*, 133.
[116] C. L. White, R. A. Padgett, C. T. Liu, S. M. Yalisove, *Scripta Metall. Mater.* **1984,** *18*, 1417.
[117] D. D. Sieloff, S. S. Brenner, M. G. Burke, *Mater. Res. Soc. Symp. Proc.* **1987,** *81*, 87.
[118] I. Baker, E. M. Schulson, J. R. Michael, *Phil. Mag. B* **1988,** *57*, 379.
[119] J. J. Kruisman, V. Vitek, J. Th. M. De Hosson, *Acta Metall. Mater.* **1988,** *36*, 2729.
[120] R. A. D. MacKenzie, S. L. Sass, *Scripta Metall. Mater.* **1988,** *22*, 1807.
[121] I. Baker, E. M. Schulson, *Scripta Metall. Mater.* **1989,** *23*, 1883.
[122] I. Baker, B. Huang, E. M. Schulson, *Acta Metall. Mater.* **1988,** *36*, 493.
[123] M. L. Jokl, V. Vitek, C. J. McMahon, P. Burgers, *Acta Metall. Mater.* **1989,** *37*, 87.
[124] O. H. Kriege, J. M. Baris, *Trans. ASM* **1969,** *62*, 195.
[125] P. Beardmore, R. G. Davies, T. L. Johnston, *Trans. Met. Soc. AIME* **1969,** *245*, 1537.
[126] S. Takeuchi, E. Kuramoto, *Acta Met.* **1973,** *21*, 415.
[127] F. E. Heredia, D. P. Pope, *Acta Met.* **1986,** *34*, 279.
[128] S. M. Ohr, *Mater. Sci. Eng.* **1985,** *72*, 1.
[129] B. Reppich, *Acta Metall.* **1982,** *30*, 87.
[130] N. Thompson, *Proc. Phys. Soc. B* **1953,** *66*, 481.
[131] J. Weertman, *Phil. Mag.* **1963,** *8*, 967.
[132] B. H. Kear, A. F. Giamei, G. R. Levrant, J. Oblak, *Scripta Metall.* **1969,** *3*, 123.
[133] B. H. Kear, A. F. Giamei, J. M. Silcock, R. K. Ham, *Scripta Metall.* **1968,** *2*, 287.
[134] B. H. Kear, A. F. Giamei, G. R. Leverant, J. M. Oblak, *Scripta Metall.* **1969,** *3*, 455.
[135] P. R. Strutt, R. A. Dodd, *Proc. 3rd Bolton Landing Conf.* (Eds.: B. H. Kear, C. T. Sims, N. S. Stoloff, W. H. Westbrook) **1969,** p. 475.
[136] B. H. Kear, G. R. Leverant, J. M. Oblak, *Trans. ASM* **1969,** *62*, 639.
[137] H. Föll, C. B. Carter, M. Wilkins, *Phys. Status Solidi A* **1980,** *58*, 393.
[138] V. Paidar, Y. Yamaguchi, D. P. Pope, V. Vitek, *Phil. Mag. A* **1982,** *45*, 883.
[139] P. Veyssière, J. Douin, P. Beauchamp, *Phil. Mag. A* **1985,** *51*, 469.
[140] A. J. Huis in 't Veld, G. Boom, P. M. Bronsveld, J. Th. M. De Hosson, *Scripta Metall.* **1985,** *19*, 105.
[141] P. Veyssière, D. L. Guan, J. Rabier, *Phil. Mag. A* **1984,** *49*, 45.
[142] P. Veyssière, *MRS Synposium* **1989,** *133*, 175.
[143] P. M. Hazzledine, Y. Q. Sun, *MRS Symposium* **1991,** *213*, 209.
[144] P. B. Hirsch, *Prog. Mater. Sci.* **1992,** *36*, 63.
[145] G. Saada, P. Veyssière, *Phil. Mag.* **1992,** *66*, 1081.
[146] D. Caillard, N. Clement, A. Couret, P. Lours, A. Coujou, *Phil. Mag. Lett.* **1988,** *58*, 263.
[147] Y. Q. Sun, P. M. Hazzledine, *Phil. Mag.* **1988,** *A58*, 603.
[148] A. Korner, *Phil. Mag.* **1988,** *A58*, 507.
[149] M. J. Mills, D. C. Chrzan, *Acta Metall. Mater.* **1992,** *40*, 3051.
[150] V. Paidar, D. P. Pope, V. Vitek, *Acta Metall.* **1984,** *32*, 435.
[151] S. Takeuchi, E. Kuramoto, *Acta Metall.* **1973,** *21*, 415.
[152] H. Saka, T. Imura, *J. Phys. Soc. Jpn.* **1969,** *26*, 1327.
[153] T. Tabata, Y. Nakajima, H. Fujita, *Jpn. J. Appl. Phys.* **1977,** *16*, 2011.
[154] R. Valle, J. L. Martin, *Electron Microsc.* **1974,** *1*, 180.
[155] P. Thomas, F. Louchet, L. Kubin, B. Jouffrey, *Jpn. J. Appl. Phys.* **1977,** *16*, 176.
[156] J. Pelissier, J. J. Lopez, P. Debrenne, *Electron Microsc.* **1980,** *4*, 30.

[157] U. Messerschmidt, M. Bartsch, *Ultramicroscopy* **1994,** *56*, 163.
[158] F. Louchet, J. Pelissier, M. Audier, B. Doisneau, *Electron Microsc.* **1994,** *1*, 271.
[159] F. Zernike, *Z. Tecn. Phys.* **1935,** *16*, 454; F. Zernike, *Phys. Z.* **1935,** *36*, 848.
[160] D. van Dijk in *Diffraction and Imaging Techniques in Material Science* (Eds.: S. Amelinckx, R. Gevers, J. van Landuyt), 2nd edn., North-Holland, Amsterdam **1978,** Vol. 1, p. 355.
[161] J. Spence, *Experimental High-Resolution Electron Microscopy*, 2nd edn., Oxford University Press, New York **1988.**
[162] P. Buseck, J. Cowley, L. Eyring, *High-Resolution Transmission Electron Microscopy*, Oxford University Press, New York **1988.**
[163] O. Scherzer, *J. Appl. Phys.* **1949,** *20*, 20.
[164] W. Coene, A. Janssen, M. Op de Beeck, D. van Dijk, *Phys. Rev. Lett.* **1992,** *69*, 3743.
[165] H. Lichte, *Adv. Opt. Electron Microsc.* **1991,** *12*, 25.
[166] G. van Tendeloo, S. Amelinckx, *Phys. Status Solidi* **1982,** *71*, 185; *ibid.* **1982,** *69*, 103.
[167] D. Schryvers, S. Amelinckx, *Metr. Ress. Bull.* **1985,** *20*, 367.
[168] D. Schryvers, J. van Landuyt, G. van Tendeloo, S. Amelinckx, *Phys. Status Solidi* **1982,** *71*, K9.
[169] D. Schryvers, G. van Tendeloo, S. Amelinckx, *Phys. Status Solidi* **1985,** *87*, 401.
[170] F. C. Lovery, G. van Tendeloo, J. van Landuyt, L. Delaey, S. Amelinckx, *Phys. Status Solidi* **1984,** *86*, 553.
[171] S. Amelinckx, G. van Tendeloo, J. van Landuyt, *Ultramicroscopy* **1985,** *18*, 395.
[172] K. Hiraga, D. Shindo, M. Hirabayashi, *J. Appl. Crystallogr.* **1981,** *14*, 185.
[173] M. Hirabayashi, K. Hiraga, D. Shindo, *Ultramicroscopy* **1982,** *9*, 197.
[174] D. Watanabe, O. Terasaki, in *Phase Transformations in Solids* (Ed.: T. Tsakalakos), North-Holland, Amsterdam **1984,** p. 231.
[175] D. Shindo, *Acta Crystallogr. A* **1982,** *38*, 310.
[176] M. El Azzaoui, J. M. Penisson, V. Pontikis, *Interface Sci.* **1994,** *2*, 79.
[177] L. Potez, A. Loiseau, *Interface Sci.* **1994,** *2*, 91.
[178] K. Urban, *Phys. Status Solidi A* **1979,** *56*, 157.
[179] N. Ishikawa, K. Furuya, *Ultramicroscopy* **1994,** *56*, 211.
[180] L. D. Marks, D. J. Smith, *Nature* **1983,** *303*, 316.
[181] D. J. Smith, L. D. Marks, *Ultramicroscopy* **1985,** *16*, 101.
[182] M. J Mills, M. S. Daw, S. M. Foiles, *Ultramicroscopy* **1994,** *56*, 79.
[183] J. M. Penisson, A. Bourret, *Phil. Mag. A* **1979,** *40*, 811.
[184] A. de Crecy, A. Bourret, S. Naka, A. Lasalmonie, *Phil. Mag. A* **1988,** *47*, 245.
[185] G. R. Anstis, P. B. Hirsch, C. J. Humphreys, J. L. Hutchison, A. Ourmazd in *Microscopy of Semiconducting Materials* (Ed.: A. G. Cullis), The Institute of Physics, Bristol **1981,** p. 15.
[186] A. Bourret, J. Thibault-Desseaux, C. D'Anterroches, J. M. Penisson, A. DeCrecy, *J. Microsc.* **1983,** *129*, 337.
[187] A. Olsen, J. C. H. Spence, *Phil. Mag. A* **1981,** *43*, 945.
[188] W. Mader, G. Necker, S. Babcock, R. W. Balluffi, *Scripta Metall.* **1987,** *21*, 555.
[189] J. M. Penisson, A. Bourret, *Phil. Mag. A* **1979,** *40*, 811.
[190] J. M. Penisson, R. Gronsky, J. B. Brosse, *Scripta Metall.* **1982,** *16*, 1239.
[191] Y. Ishida, H. Ichinose, M. Mori, M. Hashimoto, *Trans. Jpn. Inst. Met.* **1983,** *24*, 349.
[192] D. J. Smith, J. C. Barry in *High Resolution Transmission Electron Microscopy and Associated Techniques* (Eds.: P. Buseck, J. Cowley, L. Eyring), Oxford University Press, New York **1988,** p. 477.
[193] A. P. Sutton, R. W. Balluffi, *Interfaces in Crystalline Materials*, Clarendon Press, Oxford **1995.**
[194] M. I. Buckett, K. L. Merkle, *Ultramicroscopy* **1994,** *56*, 71.
[195] C. J. D. Hetherington, U. Dahmen, M. A. O'Keefe, R. Kilaas, J. Turner, K. H. Westmacott, M. J. Mills, V. Vitek, *Mater. Res. Soc. Symp.* **1990,** *183*, 45.
[196] U. Dahmen, J. Douin, C. J. D. Hetherington, K. H. Westmacott, *Mater. Res. Soc. Symp.* **1989,** *139*, 87.
[197] J. Howe, U. Dahmen, R. Gronsky, *Phil. Mag. A* **1987,** *56*, 31.
[198] D. E. Luzzi, Min Yan, M. Sob, V. Vitek, *Phys. Rev. Lett.* **1991,** *67*, 1894.
[199] T. G. Ference, R. W. Balluffi, *Scripta Metall.* **1988,** *22*, 1929.
[200] R. W. Fonda, D. E. Luzzi, *Mater. Res. Soc. Symp.* **1993,** *295*, 161.
[201] F. Phillipp, R. Höschen, M. Osaki, G. Möbius, M. Rühle, *Ultramicroscopy* **1994,** *56*, 1.
[202] G. Thomas in *Diffraction and Imaging Techniques in Material Science* (Eds.: S. Amelinckx, R. Gevers, J. van Landuyt), 2nd edn., North-Holland, Amsterdam **1978,** Vol. 1, p. 217.
[203] C. M. Wayman in *Diffraction and Imaging Technqiues in Material Science* (Eds.: S. Amelinckx, R. Gevers, J. van Landuyt), 2nd edn., North-Holland, Amsterdam **1978,** Vol. 1, p. 251.
[204] J. M. Howe, H. I. Aaronson, R. Gronsky, *Acta Metall.* **1985,** *33*, 639, 649.

[205] G. B. Olsen, M. Cohen in *Dislocations in Solids* (Ed.: F. R. N. Nabarro), North-Holland, Amsterdam **1986,** Vol. 7, p. 295.
[206] L. P. Kubin, G. Canova, M. Condat, B. Devincre, V. Pontikis, Y. Brechet, *Solid State Phenomena* **1992,** *23–24*, 455.
[207] L. P. Kubin, G. Canova in *Electron Microscopy and Fracture Research* (Eds.: U. Messerschmidt et al.), Academie Verlag, Berlin **1990,** p. 23.
[208] E. van der Giessen, A. Needleman, *ASME* **1994,** *294*, 53.
[209] J. Kratochvil, M. Saxlová, *Solid State Phenomena* **1992,** *23/24*, 369.
[210] J. Kratochvíl, A. Orlová, *Phil. Mag.* **1990,** *61*, 281.
[211] J. Kratochvíl, *J. Mech. Behav. Met.* **1990,** *2*, 353.
[212] E. C. Aifantis in *Dislocations in Solids* (Eds.: H. Suzuki, T. Ninomiya, K. Sumino, S. Takeuchi), Tokyo University Press, Tokyo **1985,** p. 41.
[213] E. C. Aifantis, *Int. J. Plasticity* **1987,** *3*, 211.
[214] E. C. Aifantis in *Non Linear Phenomena in Material Science* (Eds.: L. Kubin, G. Martin), Trans. Tech. Publications, Aedermannsdorf, Switzerland **1988,** p. 397.
[215] R. J. Asaro, *Acta Metall.* **1979,** *27*, 445.
[216] P. Hähner, L. P. Kubin, *Solid State Phenomena* **1992,** *23/24*, 385.
[217] J. Th. M. De Hosson, O. Kanert, A. W. Sleeswijk, in *Dislocations in Solids* (Ed.: F. R. N. Nabarro), North-Holland, Amsterdam **1983,** Vol. 6, p. 441.
[218] C. S. Nichols, D. R. Clarke, *Acta Metall.* **1991,** *39*, 995.
[219] C. S. Nichols, R. F. Cook, D. R. Clarke, D. A. Smith, *Acta Metall.* **1991,** *39*, 1657.
[220] C. S. Nichols, R. F. Cook, D. R. Clarke, D. A. Smith, *Acta Metall.* **1991,** *39*, 1666.
[221] D. A. Smith in *Interfaces* (Eds.: S. Ranganathan, C. S. Pande, B. B. Rath, D. A. Smith), Trans Tech. Publications, Aedemannsdorf, Switzerland **1993,** p. 87.
[222] P. J. Goodhew in *Materials Metrology and Standards for Engineering Design* (Ed.: B. F. Dyson), Elsevier, Amsterdam **1992.**
[223] P. J. Goodhew, *Electron Microsc.* **1992,** *II*, 211.
[224] J. Chone, *Int. Symp. Quantitative Metallography, Florence,* Association Italiana di Metallurgica, Milano **1978,** p. 209.
[225] M. S. Laws, P. J. Goodhew, *Acta Metall.* **1991,** *39*, 1525.
[226] J. W. Christian, *Transformations in Metals and Alloys*, Pergamon Press, Oxford **1965.**
[227] A. G. Khachaturyan, *Theory of Structural Transformations in Solids*, Wiley, New York **1983.**
[228] P. Haasen (Ed.), *Mat. Sci. Technol.* **1991,** *5*.
[229] D. A. Porter, K. E. Easterling, *Phase Transformations in Metals and Alloys*, VNR International, Berkshire, England **1988.**
[230] R. Smith, J. S. Bowles, *Acta Metall.* **1960,** *8*, 405.
[231] C. Leroux, A. Loiseau, D. Broddin, G. Van Tendeloo, *Phil. Mag. B* **1991,** *64*, 57.
[232] Z. Nishiyama, *Martensitic Transformations* (Eds.: M. Fine, M. Meshii, C. Wayman), Academic Press, New York **1978.**
[233] D. de Fontaine, *Met. Trans. A* **1988,** *19*, 169.
[234] H. Warlimont, L. Delaey, *Martensitic Transformations in Cu-, Ag- and Au-Based Alloys*, Pergamon Press, Oxford **1974.**
[235] G. Van Tendeloo, S. Amelinckx, *Acta Cryst. A* **1974,** *30*, 421.
[236] J. W. Cahn, G. Kalonji in *Solid Phase Transformations* (Eds.: H. I. Aaronson, D. E. Laughlin, R. F. Sekerka, C. M. Wayman), The Metallurgical Society of AIME, New York **1982.**
[237] D. Gratias, R. Portier in *Microscopie Electronique en Science des Matériaux* (Eds.: B. Jouffrey, A. Bourret, C. Colliex), CNRS, Paris **1983,** p. 229.
[238] W. Coene, D. Van Dyck, J. Van Landuyt, S. Amelinckx, *Phil. Mag. B* **1987,** *56*, 415.
[239] S. Amelinckx, D. Van Dyck in *Electron Diffraction Techniques* (Ed.: J. M. Cowley), Oxford University Press, Oxford **1993.**
[240] D. Van Dyck, G. Van Tendeloo, S. Amelinckx, *Ultramicroscopy* **1982,** *10*, 263.
[241] G. Van Tendeloo, S. Amelinckx, *Phys. Status Solidi A* **1978,** *49*, 337.
[242] G. Van Tendeloo, S. Amelinckx, *Phys. Status Solidi A* **1977,** *43*, 553.
[243] D. Schryvers, J. Van Landuyt, G. Van Tendeloo, S. Amelinckx, *Phys. Status Solidi A* **1983,** *76*, 575.
[244] D. Schryvers, S. Amelinckx, *Res. Mech.* **1987,** *22*, 101.
[245] G. Van Tendeloo, S. Amelinckx, *Phys. Status Solidi A* **1981,** *65*, 431.
[246] S. Amelinckx, J. Van Landuyt in *Diffraction and Imaging Techniques in Material Science* (Eds.: S. Amelinckx, R. Gevers, J. Van Landuyt), North-Holland, Amsterdam **1978,** p. 107.
[247] E. Ruedl, P. Delavignette, S. Amelinckx, *Phys. Status Solidi* **1968,** *28*, 305.
[248] D. Broddin, G. Van Tendeloo, J. Van Landuyt, S. Amelinckx, A. Loiseau, *Phil. Mag. B* **1988,** *57*, 31.
[249] D. Broddin, C. Leroux, G. Van Tendeloo in *HREM of Defects in Materials* (Eds.: R. Sinclair, D. J. Smith, U. Dahmen), MRS Society, Pittsburgh **1990,** p. 27.

[250] J. E. Spruiell, E. E. Stansbury, *J. Phys. Chem. Solids* **1965,** *26*, 811.
[251] A. G. Khachaturyan, *Phys. Status Solidi B* **1973,** *60*, 9.
[252] G. Van Tendeloo, S. Amelinckx, D. de Fontaine, *Acta Crystallogr. B* **1985,** *41*, 281.
[253] R. De Ridder, G. Van Tendeloo, S. Amelinckx, *Acta Crystallogr. A* **1976,** *32*, 216.
[254] R. De Ridder, G. Van Tendeloo, D. Van Dyck, S. Amelinckx, *Phys. Status Solidi A* **1976,** *38*, 663; *ibid.* **1977,** *40*, 669; *ibid.* **1977,** *43*, 541.
[255] P. De Meulenaere, G. Van Tendeloo, J. Van Landuyt, D. Van Dyck, *Ultramicroscopy* **1995,** *60*, 265.
[256] R. W. Cahn, P. A. Siemers, E. L. Hall, *Acta Metall.* **1987,** *35*, 2753.
[257] J. M. Sanchez, S. Eng, Y. P. Wu, J. K. Tien, *Mater. Res. Soc. Symp. Proc.* **1987,** *81*, 57.
[258] C. Leroux, A. Loiseau, M. C. Cadeville, D. Broddin, G. Van Tendeloo, *J. Phys. C: Cond. Matter* **1990,** *2*, 3479.
[259] C. Ricolleau, A. Loiseau, F. Ducastelle, *Phase Transitions* **1991,** *30*, 243.
[260] A. Finel, V. Mazauric, F. Ducastelle, *Phys. Rev. Lett.* **1990,** *65*, 1016.
[261] G. Van Tendeloo, D. Schryvers, L. E. Tanner, D. Broddin, C. Ricoleau, A. Loiseau, *Structural and Phase Stability of Alloys*, Plenum Press, New York **1992.**
[262] R. D. James, Private communication.
[263] K. Bhattacharya, *Acta Metall. Mater.* **1991,** *39*, 2431.
[264] J. M. Ball, R. D. James, *Arch. Ration. Mech. Anal.* **1992,** *100*, 13.
[265] W. A. Maxwell, E. M. Grala, *NACA TN* **1954,** *3259.*
[266] S. Chakravorty, C. M. Wayman, *Metall. Trans. A* **1976,** *7*, 555, 569.
[267] Y. Murakami, K. Otsuka, S. Hanada, S. Watanabe, *Mater. Sci. Eng. A* **1994,** *189*, 191.
[268] Y. Murakami, K. Otsuka, S. Hanada, S. Watanabe, *Mater. Trans. JIM* **1992,** *33*, 282.
[269] G. P. Mohanty, B. B. Rath, *30th Annual EMSA Meeting, Texas* **1972,** p. 584.
[270] D. Schryvers, *Phil. Mag. A* **1993,** *68*, 1017.
[271] I. Baele, G. Van Tendeloo, S. Amelinckx, *Acta Metall.* **1987,** *35*, 401.
[272] M. Chandrasekaran, K. Mukherjee, *Mater. Sci. Eng.* **1974,** *13*, 197.
[273] M. Chandrasekaran, L. Delaey, *ICOMAT, Leuven* (*J. Phys. C4*) **1982,** p. 661.
[274] F. Reynaud, *Scripta Metall.* **1977,** *11*, 765.
[275] V. V. Martynov, K. Enami, L. G. Khandros, A. V. Tkachenko, S. Nenno, *Scripta Metall.* **1983,** *17*, 1167.
[276] D. Schryvers, L. E. Tanner, *Shape Memory Materials and Phenomena*, MRS, Boston **1991,** Vol. 246, p. 33.
[277] T. Inoue, S. Morito, Y. Murakami, K. Oda, K. Otsuka, *Mater. Lett.* **1994,** *19*, 33.
[278] Y. Nakata, T. Tadaki, K. Shimizu, *Ultramicroscopy* **1991,** *39*, 72.
[279] J. Pons, L. Jordan, J. P. Morniroli, R. Portier, *J. Phys. IV* **1995,** *5*, C2.
[280] D. Cherns, A. R. Preston, *J. Electr. Microsc. Technol.* **1989,** *13*, 111.
[281] M. Tanaka, M. Terauchi, T. Kaneyana, *CBED II*, JEOL Ltd., Tokyo **1988.**
[282] P. J. Othen, M. L. Jenkins, G. D. W. Smith, *Phil. Mag. A* **1994,** *70*, 1.
[283] H. A. Hardouin Duparc, R. C. Doole, M. L. Jenkins, A. Barbu, *Phil. Mag. A* **1995,** *71*, 325.
[284] L. E. Tanner, *Phil. Mag.* **1966,** *14*, 111.
[285] I. M. Robertson, C. M. Wayman, *Phil. Mag.* **1983,** *48*, 421, 443, 629.
[286] T. Krekels, G. Van Tendeloo, D. Broddin, S. Amelinckx, L. Tanner, M. Mehbod, E. Vanlathem, R. Deltour, *Physica C* **1991,** *173*, 361.
[287] D. Schryvers, L. E. Tanner, *Ultramicroscopy* **1990,** *32*, 241.
[288] G. Van Tendeloo, S. Amelinckx, *Scripta Metall.* **1986,** *20*, 335.
[289] D. Schryvers, L. E. Tanner, *Trans. Mater. Res. Soc. Jpn.* **1994,** *18B*, 849.
[290] D. Schryvers, L. E. Tanner, *Mat. Sci. Forum* **1990,** *56–58*, 329.
[291] J. Stoiber, B. Guisolan, R. Gotthardt, *Ultramicroscopy* **1993,** *49*, 37.
[292] T. Saburi, *ICOMAT, Monterey 1992, Proceedings* (Eds.: C. Wayman, J. Perkins) **1992,** p. 857.
[293] J. M. Howe, *ICOMAT, Monterey 1992, Proceedings* (Eds.: C. Wayman, J. Perkins) **1992,** p. 185.
[294] J. M. Howe, S. J. Rozeveld, *J. Microsc. Res. Technol.* **1992,** *23*, 230.
[295] S. Muto, N. Merk, D. Schryvers, L. E. Tanner, *Phil. Mag. B* **1993,** *67*, 673.
[296] S. Muto, D. Schryvers, N. Merk, L. E. Tanner, *Acta Metall. Mater.* **1993,** *41*, 2377.
[297] D. Schryvers, Y. Ma, L. Toth, L. E. Tanner, *Acta Metall. Mater.* **1995,** *43*, 4045.
[298] D. Schryvers, Y. Ma, L. Toth, L. E. Tanner, *Acta Metall. Mater.* **1995,** *43*, 4057.
[299] T. Tadaki, C. J. Qiang, K. Shimizu, Y. Fuxing, G. Nanju, *ICOMAT, Monterey 1992, Proceedings* (Eds.: C. Wayman, J. Perkins) **1992,** p. 827.
[300] T. Tadaki, K. Shimizu, *Metall. Mater. Trans. A* **1994,** *25A*, 2569.
[301] M. A. O'Keefe, *Ultramicroscopy* **1992,** *47*, 282.

[302] F. Philipp, R. Höschen, M. Osaki, G. Möbus, M. Rühle, *Ultramicroscopy* **1994,** *56*, 1.
[303] P. J. Wilbrandt, *Ultramicroscopy* **1993,** *60*, 193.
[304] O. L. Krivanek, P. A. Stadelmann, *Ultramicroscopy* **1995,** *60*, 103.
[305] J. M. Cowley, *Acta Cryst. A* **1988,** *44*, 847.
[306] O. L. Krivanek, C. C. Ahn, G. J. Wood, *Ultramicroscopy* **1990,** *33*, 177.
[307] M. J. Hÿtch, W. M. Stobbs, *Ultramicroscopy* **1994,** *53*, 191.
[308] G. Möbus, M. Rühle, *Ultramicroscopy* **1994,** *56*, 54.
[309] K. L. Merkle, *Ultramicroscopy* **1992,** *40*, 281.
[310] G. J. Wood, W. M. Stobbs, D. J. Smith, *Phil. Mag. A* **1986,** *50*, 375.
[311] J. Mayer, G. Gutekunst, G. Möbus, J. Dura, C. P. Flynn, M. Rühle, *Acta Metall. Mater.* **1002,** *40*, S217.
[312] A. Trampert, F. Ernst, C. P. Flynn, C. P. Fischmeister, M. Rühle, *Acta Metall. Mater.* **1992,** *40*, S227.
[313] W. O. Saxton, D. J. Smith, *Ultramicroscopy* **1985,** *18*, 39.
[314] G. Möbus, G. Necker, M. Rühle, *Ultramicroscopy* **1993,** *49*, 46.
[315] D. Hofmann, F. Ernst, *Ultramicroscopy* **1994,** *53*, 205.

Handbook of Microscopy
S. Amelinckx, D. van Dyck, J. van Landuyt, G. van Tendeloo

2 Microscopy of Rocks and Minerals

2.1 Introduction

Many minerals have lower symmetry than cubic. In general, the nature of their bonding causes a large gap between the valence and conduction bands, so that the most minerals are excellent insulators and, correspondingly, transparent to visible light. Opaque minerals are either semiconducting or metallic. The particular physical properties of minerals, and the rocks that they form, dictate and somewhat limit the types of microscopy that can be applied. As a compensation, the anisotropy deriving from the noncubic symmetries of many minerals provides the basis for phase identification in optical petrology and the determination of crystal orientations.

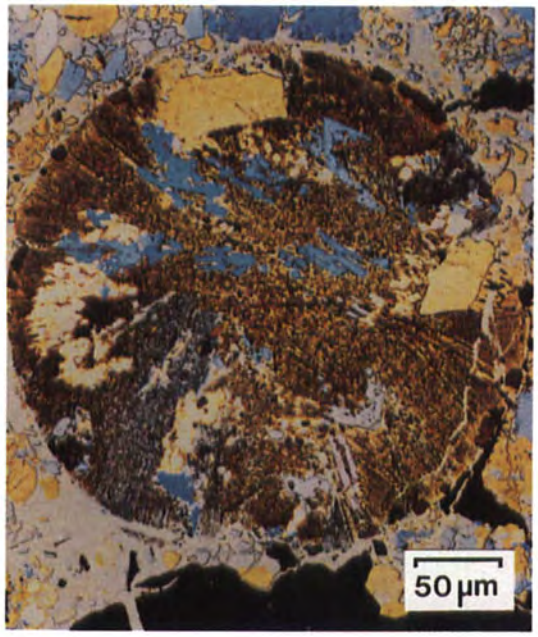
When minerals are irradiated with electrons or ions the problems of electrical charging preclude some techniques that provide abundant data about metals and semiconductors. For instance, no results have come from field ion microscopy, and Auger electron spectroscopy (AES) is severely limited in application. Spectroscopic methods in imaging mode, like Raman, FTIR and NMR, are at an early stage of application in the geosciences and there are relatively few examples. The range of spectroscopies that have been applied to problems in geoscience are discussed elsewhere [1, 2]. Apart from IR mapping and Raman imaging, only AES and XPS seem to merit a separate discussion at this stage.

In just one chapter it is impossible to discuss and reference all the excellent microscopy that has been carried out on rocks and minerals over many years. The material covered is inevitably very selective and the emphasis reflects a personal view. I have endeavored to emphasize the state of the art for the newer and less routine techniques, since the established methods are well documented elsewhere. I apologize in advance if my bias or ignorance has led to the omission of technical developments or findings that others consider to be important.

2.2 Optical Microscopy, Petrology, and Other Optical Methods

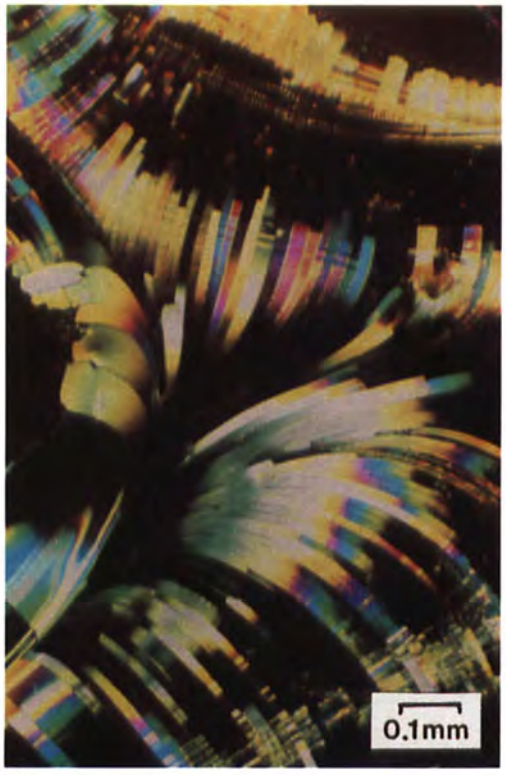
2.2.1 Transmitted Light Microscopy and Petrology

Light optical microscopy (LOM) has been a routine and essential tool of the earth scientist for many years. The theory of



(a)

(b)



(c)

(d)

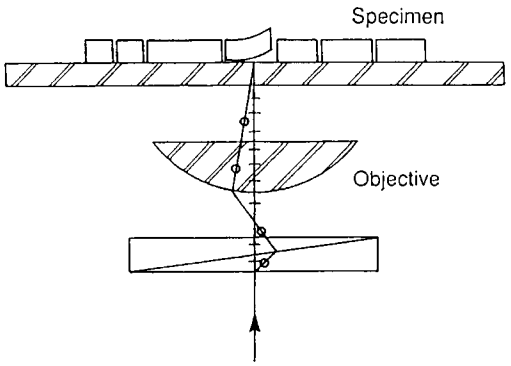


Figure 1. (e) Schematic diagram of an arrangement for producing inverted-type differential interference contrast images from thin translucent crystals, as in (c) and (d). Reproduced with permission from Bousfield [6] and reprinted by permission of John Wiley & Sons Ltd.

crystal optics enables analysis of individual crystals and mineral grains in transmission with the polarizing microscope [3]. The study of a polished thin section (PTS) with the polarizing microscope allows all the routine methods of petrography to be applied to the identification of minerals in a rock. These techniques, which include the study of the color and pleochroism of a mineral, its shape and habit (noting cleavages, twinning, relief, inclusions, etc.), followed by establishing its optical characteristics (refractive indices, optic sign, etc.), are applied routinely by earth scientists and will not be described here. Standard texts cover the theory of optical mineralogy and describe the methods of optical petrology [4].

The preparation of high quality PTSs reproducibly, is basic to effective optical petrology, which starts from cutting slices with a diamond saw, followed by lapping and polishing with progressively finer powders to a thickness of $\approx 30\,\mu m$. Techniques and materials used in the preparation of PTSs of geological materials are well covered in the literature [5, 6]. Examples of minerals and rocks in thin section are not illustrated here, since they are readily accessible [7, 8].

Ultrathin polished sections ($<30\,\mu m$ thick and usually $<10\,\mu m$) require special care in preparation [9]. They take best advantage of the resolution in LOM ($\approx 0.5\,\mu m$) and are invaluable for very fine-grained materials, since the superposition of grain-scale features is minimized. The better definition removes sources of ambiguity and gives very beautiful images [10]. Figures 1a and b show ultrathin sections of two stony meteorites in transmitted polarized light. Demountable polished thin sections (DPTSs) [5] and ultrathin versions [11] are a further development; their value is mainly as starting points for making specimens for transmission electron microscopy.

The universal stage (U-stage) and spindle stage were once essential for making precise crystallographic measurements on minerals and petrofabrics. The U-stage featured strongly in studies of carbonates [12] and deformed tectonites [13], works

Figure 1. Transmitted light images of ultrathin polished sections (2–4 μm thick) of stony meteorites viewed with partially light and quartz plate: (a) Allegan (H5) chrondritic meteorite, showing olivines (purple, blue, yellow) and chromites (brown) in a radiating fine-grained chondrule and some of the surrounding matrix; (b) Dhajala (H3) chondritic meteorite, showing a chondrule consisting almost entirely of extremely fine-grained pyroxene with additional Fe/Ni metal and troilite in the irregular core. Micrographs taken by R. Beauchamp [10]. Reprinted by permission of Meteorifics. Inverted differential interference contrast images of columnar sodium silicate crystals grown in colloidal silica solution, seen with (c) an extinguished background and (d) a red background obtained by adjusting the wedge position. 100×. Courtesy of Bousfield [6].

that are prime examples of the method. Electron microscopy, with its greater resolution and other advantages, has eclipsed the U-stage but its effectiveness at the submillimeter scale should not be overlooked [14].

The polarizing microscope has been used in texture analysis by a photometric method [15] but the method has not found favor. It is not so versatile and less attractive than automated analysis by EBSP (see this Volume, Chap. 2, Sec. 2.6.3). The photometric approach has greater merit for clays and soils, where it is quite widely used and the benefits of image analysis have been noted [16].

2.2.2 Reflected Light Microscopy and Opaque Minerals

For opaque minerals (sulfides, hematite, etc.), optical techniques are much more restricted than for transparent minerals. Color, lustre, cleavage, twins, and exsolved phases in crystallographic relationship to the host mineral are all features, visible in reflected light, that can assist in phase identification and crystallographic analysis. The reader should see texts on ore mineralogy for further information [17]. Polishing followed by etching, thermally or chemically, can be a valuable adjunct for particular minerals. Standard techniques used in the crystallographic analysis of deformed metals by LOM [18] (see also this Volume, Chap. 1, Sec. 1) are also applicable to the opaque minerals.

The use of differential interference contrast (DIC) microscopy and phase contrast (PC) microscopy is most evident in studies of mineral surfaces, and especially in relation to natural growth features, laboratory studies of crystal growth, impurity segregation, etc. Good examples of DIC and PC are given in references [19] and [20], respectively. Figures 1c and d are representative of some striking DIC images of thin minerals produced [6] using an inverted viewing arrangement, shown in Fig. 1e, which is sensitive to surface slope, surface interference, and dichroism. (For comparison, a normal DIC image of a crystal surface is shown in Sec. 2.5, Fig. 4d.)

2.2.3 Infrared Micro-spectroscopy, Laser Raman Microprobe and Micro-Raman Imaging

Infrared spectroscopy gives valuable information about organic species and elements like H, C and N which they incorporate. It can now be employed to map distributions, and to reveal zonations and inclusion chemistry in minerals. This is well illustrated in studies of diamond [21] which use mid-infrared micro-spectroscopy in combination with other techniques, including cathodoluminescence (see Sec. 2.8 of this Chapter). Software developments allow contour maps for various chemically-significant spectral lines to be generated relatively easily.

2.2.3.1 Phase Identification

Many minerals with absorption spectra in the middle and far infrared spectrum conveniently give Raman lines in the visible spectrum. Raman spectra offer an excellent analytical finger-printing technique,

the lines being relatively simple, but occurring in distinctive and unique patterns. Applications in mineralogy are assisted by a compilation of Raman spectra [22]. New instrumentation allows an image to be formed by selecting a single Raman line, providing a simple, nondestructive method of sequentially imaging different phases in a bulk sample, which may be useful when their optical characteristics are similar and where cathodoluminescence (CL) (Sec. 2.5 of this Chapter) does not work well. This micro-Raman method is also a means of detecting foreign matter that has been introduced artificially, for example, the impregnation or dyeing of inferior semiprecious minerals, to increase their value.

Micro-Raman is also very useful for in situ studies of minerals in high pressure cells. There are various examples of spectroscopy with diamond anvil cells, although, as yet, without imaging [23, 24]; it is, nevertheless, a growing field of activity.

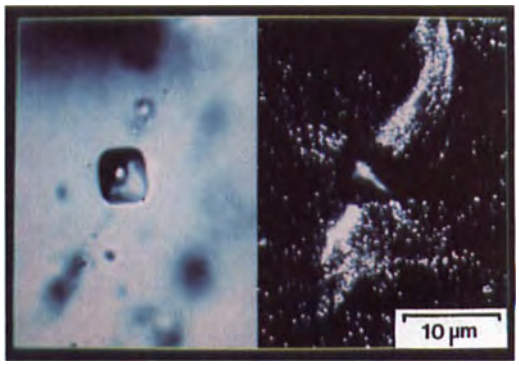


Figure 2. (a) Transmitted light image two-phase (water and gas) fluid inclusion in quartz. Scalebar: 10 µm, laser wavelength 488 nm; (b) reflected confocal scanning laser microscopy image made up of 80 optical sections taken at 0.3 µm increments, representing a depth of 24 µm. The fluid inclusion was located ≈60 µm below the surface. The confocal image reveals the existence of a large number of highly reflective, submicron inclusions not seen in the conventional image. Courtesy of Petford [27].

2.2.3.2 Fluid Inclusions

The majority of minerals form from fluids. Fluid inclusions are important because they are a record of the history of the host mineral. In principal, analysis of inclusion contents enables pressure, temperature, density, and compositional data to be established for geological fluids, present when a mineral formed or was altered. Results of fluid inclusion research from optical methods are summarized and listed by Roedder [25], a leader in the field for many years.

The use of microthermometry with LOM is common, and an excellent recent example is the combination of CL (see this Volume, Chap. 1, Sec. 6) with microscopy to obtain a fluid inclusion stratigraphy related to the authigenic minerals and cement zones within them [26]. Confocal scanning laser microscopy (CSLM) (see Sec. 2.2.4 of this Chapter) is also valuable in fluid inclusion research, since it can reveal the presence of features not seen by routine LOM [27]. Figures 2a and b illustrate this point very well.

The challenge today is to obtain quantitative analysis of a few cubic micrometers of fluid in an inclusion (often a supercritical liquid), typically containing dissolved gases and solids, the entity being isolated within a semitransparent solid. The laser Raman microprobe is an instrument well suited to this task [28–30]. The laser beam can be focused so that the contents of an inclusion can be analyzed from its spectra nondestructively and very valuable results

have been obtained. A theory for the scattering of light by fluids under pressure is still being developed, however, and calibration data for the interpretation of the resulting spectra still need to be acquired [31]. Proton-induced X-ray emission (PIXE) is another technique that has similar uses in the study of fluid inclusions [32] and can be a microscopic technique (see Sec. 2.12 of this Chapter).

2.2.4 Confocal Scanning Laser Microscopy

The confocal scanning laser microscope (CSLM) is a new tool which has so far found little application in the earth sciences, although its resolution and selectivity at depth makes it very suitable for some types of work. CSLM has been used by a few workers for improved imaging of microfractures and quantitative measurements of porosity. It has used it to characterize the fissure space of rocks [33]. Correlated studies of sections were carried out by several microscopic methods, including acoustic microscopy and fluorescence microscopy. CSLM was shown to be a superior technique for observing the fine detail of fractigraphic networks, facilitating stereology and digital image analysis.

Observations of etched fission tracks and the reliable counting of track densities are essential to the well-established fission-track method of age determination. For currently evolving methods of estimating paleotemperatures, both track lengths and track densities must be determined accurately. Conventional LOM does not work well for tracks at high angles to the plane of the section and, more generally, there are problems with distinguishing tracks from some other forms of defect. Both aspects are less problematical when CSLM is used. This is exemplified by the use of CSLM for imaging etched fission tracks [34]. The CSLM seems likely to revolutionize track analysis as software becomes available to permit the automated analysis of 3D binary images [35]. CSLM is now also being applied to fluid inclusions (see also Sec. 2.2.3 of this Chapter).

2.2.5 Ultraviolet Fluorescence Optical Microscopy

Ultraviolet fluorescence microscopy provides a good alternative to scanning acoustic microscopy or scanning electron microscopy (SEM) methods for the study of cracks and pores. Nishiyama and Kusuda [36] have demonstrated the identification of pores and microcracks in resin-impregnated sections of granite, weathered tuff, schist and marble by ultraviolet microscopy. Cyanoacrylate and methylmethacrylate are both good permeating agents since they readily penetrate pore spaces and cracks; they can be dyed with fluoresceine or loaded with fluorescent paints. Image processing and analysis techniques, respectively, (i) allow better discrimination of the resin-filled spaces, and (ii) facilitate quantitative measurements. Figures 3a and b are excellent illustrations of the method, which is an alternative to and may be a better way of quantifying cracks and pores than SEM methods [37].

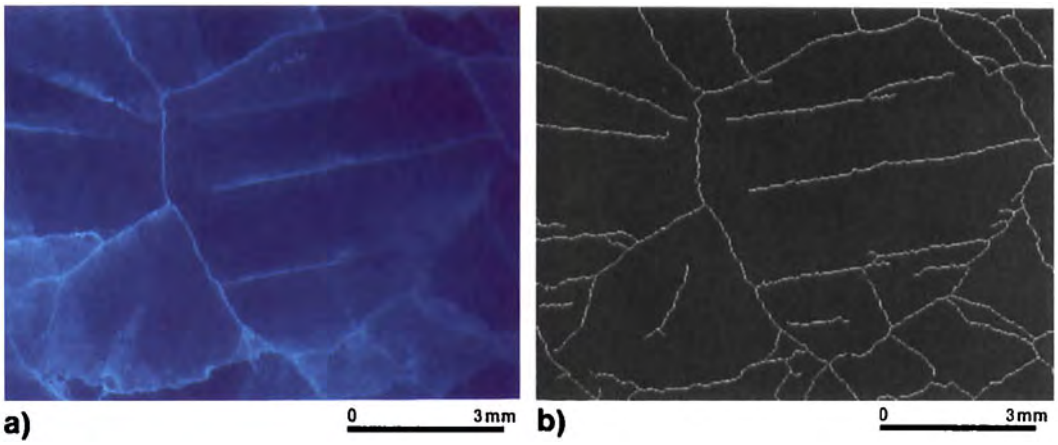


Figure 3. Ultraviolet fluorescence imaging: visualization and image analysis of microcracks in marble. (a) Visualization with fluorescent paint under ultraviolet light; (b) thinned binary image of the microcracks. Courtesy of Nishiyama and Kusuda [36].

2.3 X-Ray Methods

2.3.1 Computer-Aided Tomography

Computer-aided tomography (CT) with X-ray radiography is hardly within the scope of this book. However, it is worth noting that applications of CT in the earth sciences are increasing, although only one example will be given. This is the use of X-ray CT for investigating sediment cores, which are very important in geological prospecting. Normal X-ray radiography is of limited use with full-round cores because of defects at the core periphery, next to the sleeve. CT avoids destructive cutting and enables high resolution imaging of internal features without confusion [38]. CT also makes possible the generation of 3D images. Laser scattering CT, which can be used to detect and characterize defects near surfaces and in the bulk, is effective with hematite crystals [39] when using infrared wavelengths.

2.3.2 X-Ray Microtopography

X-ray topography has not been applied widely to minerals because of its inability to image very defective crystals. Most work has been done on diamond and quartz; this is described in this Volume, Chap. 1, Sec. 7. Both X-ray topographs and surface microtopographs have been used to study twinning and related impurity effects in hematite crystals [40].

Synchrotron radiation is being used increasingly for topographic studies of diamond and quartz. It has been used in conjunction with infrared absorption microscopy and CL topography to study the long-range imperfection of diamond [41, 42]. Greater use of in situ synchrotron X-ray radiation topography is also likely. For example, it has been employed to study

annealing processes in a sanidine feldspar with an unusually low Al/Si order–disorder transformation temperature [43].

2.4 Scanning Acoustic Microscopy

The technique of scanning acoustic microscopy is yet to be applied to rocks in a systematic way. Exploratory results are reported for granite [44] and comparisons of images of cracks in polished rock sections have been made with those from other imaging methods [45]. Continuing developments with time-resolved imaging [46] hold some promise for the study of microfracturing dynamics in laboratory studies of rock deformation.

2.5 Cathodoluminescence Microscopy

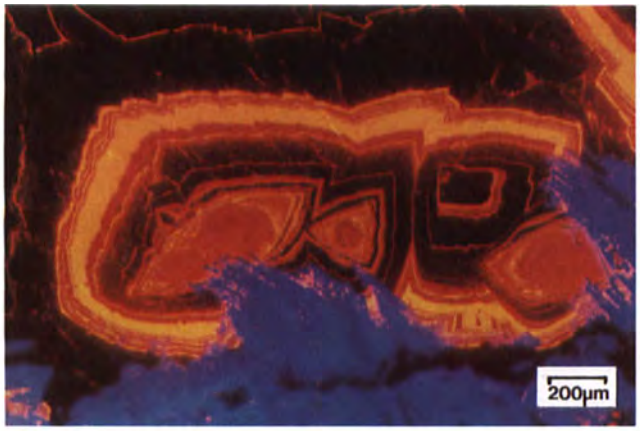
Few minerals are naturally photoluminescent, since their band gaps exceed 5 eV. Most respond to higher energy electron irradiation, luminescing with characteristic colors. This cathodoluminescence (CL) is very convenient for revealing the multiphase nature of rocks and natural cements, and compositional zoning in mineral grains and gemstones. CL is usually obtained using polished specimens, either in a dedicated instrument or an SEM.

The CL of geological materials, the influence of defects, temperature, etc., together with applications, are well described [47, 48]. CL reveals spatial distributions of minerals, distinguishing constituents with similar optical properties, for example, calcite (yellow–orange) and dolomite (crimson). Another strength is the display of significant textural features not seen by any other technique [49]. CL, however, is largely a qualitative analytical tool, since intensity and absolute concentration of the activated species are not well correlated.

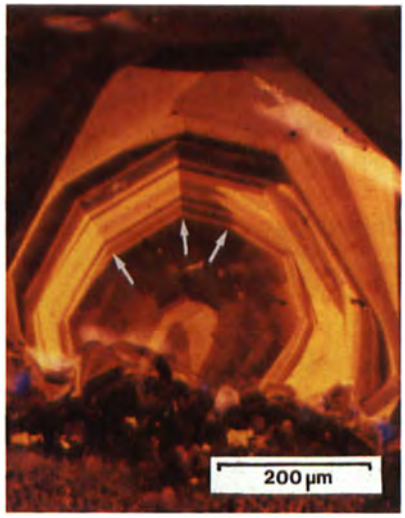
Carbonates luminesce brightly and stably. Growth features that are difficult to see by LOM are readily resolved: zoning, fast and slow-growing sectors, veins, and authigenic overgrowths. Differing trace levels of divalent ions (Mg, Fe and Mn principally) reveal successive layers of void-filling cements in carbonate rocks. Figure 4a shows a CL image of zoned calcite crystals cementing a cavity formed by dissolution of a gypsum nodule [50], while in Fig. 4b CL reveals growth sector boundaries in another cementatious calcite crystal [51]. Figures 4c and d are CL and DIC images respectively of the (0 0 0 1) surface of an apatite crystal [52].

CL is especially valuable for provenance studies, since grain populations usually possess defining spectral 'fingerprints' deriving from their source rocks, whether carbonates or sandstones. Quartz minerals are damaged by electron irradiation; their luminescence is not stable and color changes are rapid. Authigenic overgrowths do not normally luminescence but hydrothermally-grown quartz shows zoning in CL, and so it is a sensitive indicator of growth processes.

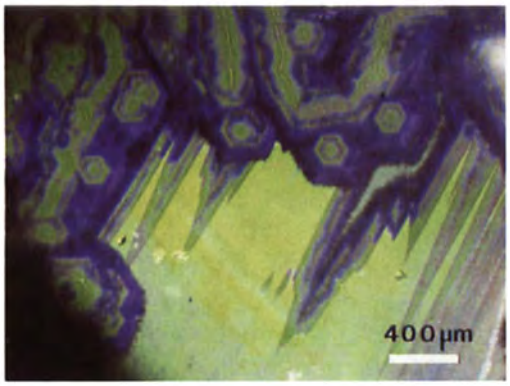
Steele [53, 54] has used CL effectively with meteoritic minerals. Only some olivines in carbonaceous chondrites show CL, but for these it is very sensitive to composition, allowing classification into subgroups and defining stages in the



(a)



(b)



(c)

(d)

Figure 4. (a) CL image of zoned calcite crystals cementing a cavity formed by dissolution of a gypsum nodule within the Upper Permian Raisby Formation at an outcrop in north-east England. Prior to dissolution of gypsum, the nodule was partially replaced by fluorite, which luminesces blue. Long dimension is 2.25 mm. Courtesy of Lee. (b) CL image of a polished section through a calcite crystal occurring as a cement in a limestone. The crystal, having formed on a substrate (at bottom), exhibits well-developed concentric growth zoning that records the crystal habit. The luminescence is of Mn^{2+} activated, and the intensity reflects subtle variations in Mn^{2+} concentration introduced during growth. Growth on different faces resulted in differential Mn^{2+} incorporation (i.e., sector zoning). Sector boundaries are arrowed. Courtesy of Reeder [53]. (c) CL image of (0001) surface of apatite crystal (from Llallagua, Bolivia) with hexagonal growth hillocks. Differential luminescence corresponds directly to symmetrically nonequivalent regions of the hillocks. (d) DIC image of the (0001) surface of apatite shown in (c) at a slightly different magnification. Both (c) and (d) courtesy of Rakovan and Reeder [54].

formation of grains with complex textures and chemistries. Olivines with blue CL were deduced to be products of nebular condensation, whereas subsequent fluid-phase reactions create red CL.

Recent developments include an instrument designed for producing strong CL in minerals of low luminescent intensity [55]. CL can be used to image defects in minerals (e.g., [56]), although this is less common than with semiconductors. Improvements in defect imaging involve computer processing of CL maps [57].

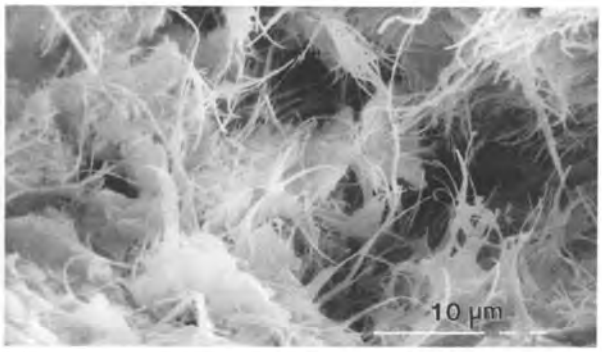
Induced thermoluminescent emissions (TL) are used in archeological and meteorite research to characterize devitrification and low-grade metamorphism. For example, TL from feldspar derived by devitrification of glassy phases is a sensitive function of petrologic grade [58]. Microscopic imaging by means of TL is a recent development [59]; two-dimensional photon counting is achieved with a highly parallel processor array system. The formation of images by triboluminescence (emission caused by fracture or wear) appears to be possible with similar equipment.

2.6 Scanning Electron Microscopy

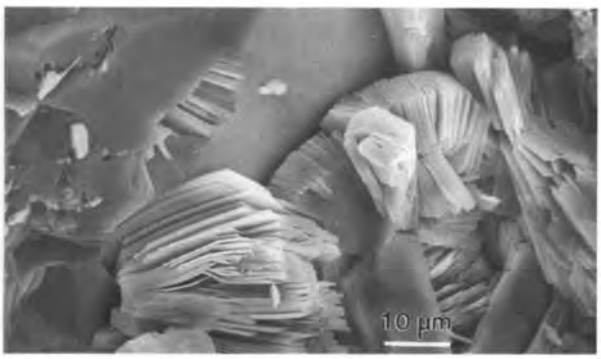
2.6.1 General Uses

The SEM is an important tool in the earth sciences. The large depth of field of the SEM makes it particularly effective in sedimentology, where fracture surfaces are used to elucidate diagenesis by examining rock texture, porosity, cements, and pore-filling materials. Stereo pairs are a useful adjunct and the heterogeneity of many sedimentary rocks favors the making of montages of images. Grains can be analyzed and identified in situ using EDS. Standard polishing methods are applicable when preparing SEM specimens of less porous rocks. All modes of SEM operation are then possible. Unconsolidated sediments can be infiltrated with resin, or prepared as scattered grain mounts and polished. Etching may assist to reveal microstructures and in deriving grain origins; for example, eroded grain surfaces may allow their transport history to be deduced. In particular, the SEM plays a pivotal role in the study of carbonate sediments, where porosity is not only characteristic of processes (e.g., dolomitization), but has important commercial implications for extraction from oil reservoirs. Figure 5 illustrates the use of the SEM in imaging to various forms of sedimentary materials [60].

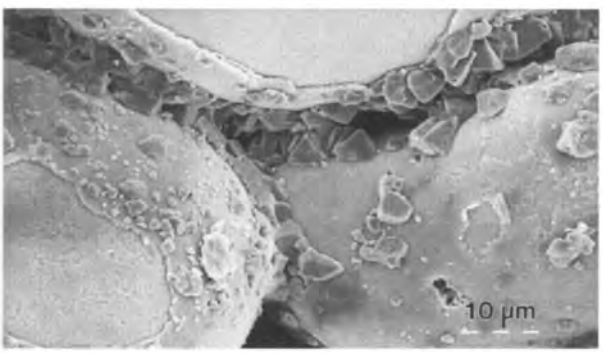
Conditions in the SEM are less harsh than in TEM and even better in an environmental SEM (ESEM), where greater stability is traded for some loss of resolution. Thus, the SEM can give excellent morphological information of radiation-sensitive minerals. When clay minerals or wet sediments are involved, it is desirable to adopt special preparation methods that do not alter the grain morphologies. These methods include freeze-drying and critical point drying (CPD), the latter being slow but preferable, because of the removal of the damaging effects of surface tension. CPD is better for the preservation of delicate filamentary structures [61]. Smart and Tovey [62, 63] illustrate the value of SEM in clay mineralogy; many excellent micrographs are given in reference [62], and the techniques are described in reference [63].



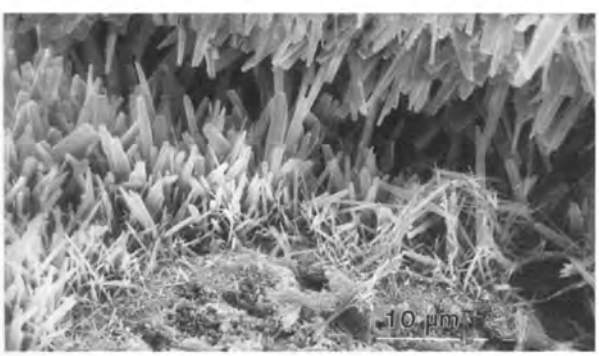
(a)



(b)



(c)



(d)

Figure 5. SEM secondary electron images of sedimentary materials. (a) Filamentous illite growing into pore space. Rotliegend aeolian sandstone, Lower Permian, southern North Sea; (b) Books of coarse kaolinite in pore space. Well developed quartz over-growths partly post-date kaolinite formation as shown by 'impressions' of kaolinite in quartz. Mungaroo Formation, Flinders Shoal No. 1, N.W. Australian Shelf; (c) Blocky calcite cement concentrated at grain contacts as a meniscus cement in an oolite cemented in the vadose zone. Lithified sub-recent oolitic aeolian dune deposit. Carbla Point, Hamelin Pool, Shark Bay, Western Australia; (d) Bladed aragonite cement. Recent submarine cementation of oolitic hardround in Hamelin Pool, Shark Bay, Western Australia. Courtesy of N. Trewin [60] and by permission of Blackwell Science Ltd.

A more specific example of the power of SEM in clay mineralogy is given by Keller and Haenni [64], on the intimate mixing of kaolins.

Direct studies of cracks and pore spaces are somewhat under-represented in the earth sciences, reflecting the difficulty of obtaining three-dimensional and quantitative data. Indirect methods of investigating porosity are easier to use, but the data can be a cloak for a wide range of features whose individual identities and significance are hidden. SEM is useful for the study of cracks and similar extended features because of the instrument's large depth of focus; it performs well at a scale intermediate to LOM and TEM. It is also excellent for studying fracture surfaces but the recognition of cracks in such surfaces which penetrate into the bulk can be difficult. Flat, polished surfaces are better for seeing cracks and pores, but polishing may obscure detail and lead to erroneous results because of infilling and the distortion of surface layers. The ion bombardment of surfaces to erode damage produced by polishing is very effective [65]. It is an important, almost essential, approach when SEM is used for crack studies. Good examples exist of SEM in the study of micropores [66] and of surface channels on grains undergoing pressure solution, Fig. 6 [67].

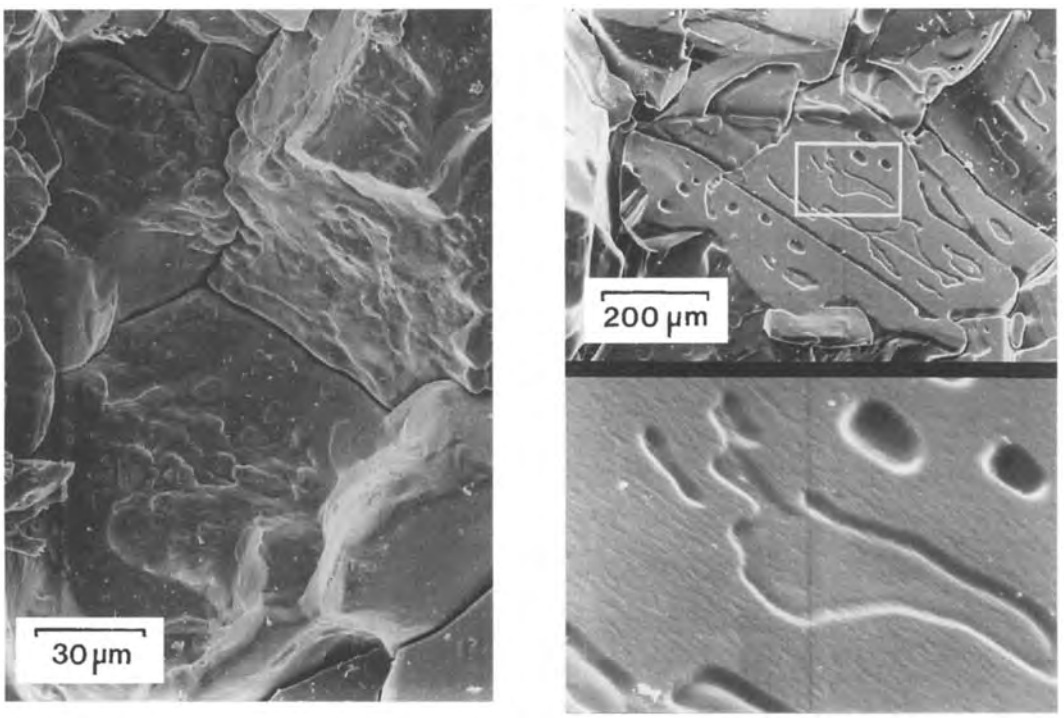


Figure 6. SEM secondary electron images showing two different forms of grain contact morphology on mechanically-parted crystals from wet-compacted samples of halite aggregate subjected to a volume strain of >25%, and undergoing pressure solution. (a) Irregularly-shaped tabular plateaus, or 'islands'; (b) (top) surfaces with an anastomozing pattern of channels, incizing into a more or less planar surface, arising from tubular structures; (bottom) enlargement of area outlined by white rectangle. Courtesy of Schutjens [67]. Figure 6(b) is reproduced by permission of the Geoligical Society, UK.

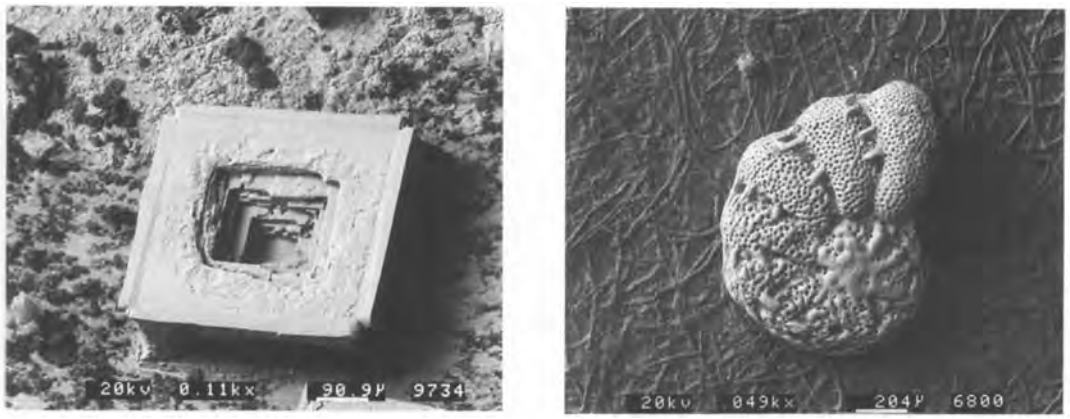


Figure 7. SEM secondary electron images (20 kV) of uncoated specimens obtained with an environmental SEM fitted with a charge-free anticontamination system: (a) halite hopper crystal, bar mark 91 µm; (b) modern forminifera from N.E. Australia, bar mark 204 µm. The specimen in (b), which was mounted on cardboard, belongs to the Micropaleontology Collection, Natural History Museum (NHM), London. Courtesy of Jones, NHM.

ESEM (see Vol. 1, Chap. 4, Sec. 1), is still too new to have been applied widely, but it clearly has advantages in approaching the humidity conditions experienced by many natural samples. High quality imaging of uncoated, nonconducting specimens is possible, through the ionization of residual air in the chamber. The usefulness of ESEM is illustrated by Figs. 7a and b, which show a hopper crystal of halite (NaCl) and a foraminifer, respectively. The latter sample was imaged undisturbed on its cardboard museum display tray (it is good policy not to coat museum-type materials, because this precludes further study by LOM).

2.6.2 Electron Channeling Patterns

Electron channeling patterns (ECP) enable crystallographic orientation to be determined (see Vol. 1 of this Handbook). The method has been explored for the petrofabric analysis of quartzites [68]. More general applications to geological materials have also been discussed [69]. The ECP technique has also been applied to twinned quartz crystals [70], showing that left and right-handed Brazil twins cannot be distinguished because the twin plane coincides with the mirror plane in the ECP sphere (i.e., the twins have identical ECPs). Dauphiné twins are readily distinguished. The ECP technique is not widely used and it is being superseded by the following methodology.

2.6.3 Electron Backscattering Patterns

The use of electron backscattering diffraction patterns (EBSPs) in conjunction with SEM is a comparatively new technique for earth scientists. Like ECP it does not image the material in an exact sense, but it gives an 'orientation image'. The EBSP technique merits a brief mention on account of its growing importance.

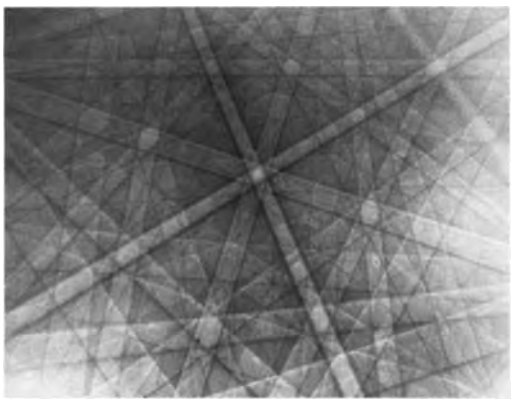


Figure 8. Wide angle backscattered electron Kikuchi diffraction pattern from chalcopyrite, illustrating the projection symmetry of the [0 0 1] axis, which has Laue group 2m. The pattern coverage along the diagonal is 70°. Courtesy of Baba-Kishi, Hong Kong Polytechnic University.

Because backscattered electrons only sample a very thin layer at the surface, it is necessary to use special specimen preparation techniques to prevent lattice distortion and 'smearing' of the surface. EBSPs are very useful for examining the symmetry of single crystals and individual mineral grains [71]. Figure 8 is the pattern for pyrite with a ⟨1 1 1⟩ pole at the center. An atlas of many patterns exists [72], including some for minerals.

More generally EBSPs are very valuable for the determination of preferred orientation, or texture, in combination with microstructural data. They also offer a means of studying plastic deformation and dislocation densities, but this approach has not yet been used for rocks. Current developments in instrumentation are much concerned with facilitating unambiguous orientational analysis by means of EBSP, and enabling orientations and textures to be measured automatically. Good examples of the power of automated systems are [73] on quartzite and [74] on calcite.

2.6.4 Microanalysis by Scanning Electron Microscopy and Electron Probe Microanalysis

Microanalysis with SEM, energy dispersive spectroscopy (EDS) and wavelength dispersive spectroscopy (WDS) (i.e., SEM/EDS, SEM/WDS) and the electron probe microanalyzer (EPMA) are essential tools of the earth scientist, generating detailed compositional data at a submillimeter level about suites of rock and minerals. Established texts and Vol. 2 of this Handbook discuss the particular problems of microanalysis and just a few points are made here.

The chemical complexity and the inhomogeneity of many geological samples outweighs that of the average metallurgical specimen. This poses particular problems when coupled with the difficulty of analyzing oxygen and other light elements quantitatively. Ideally, a complete analysis should list the phases present, the average and spread of compositions for each phase, the phases present, the average and spread of compositions for each phase, the amount of each phase, and the bulk composition of the sample. As described in Vol. 2 of this Handbook, ZAF-type computer programs are widely available to process raw WDS X-ray data so that accurate quantitative analyses result. Special programs exist for geological materials, of which the Bence–Albee correction procedure [75] for EPMA data from oxides and silicates is well known. With some modifications it is still one of the best, being rapid and easy to apply [76]. SEM/EDS is able to provide good analytical information about rock samples if data is treated by multivariate statistical

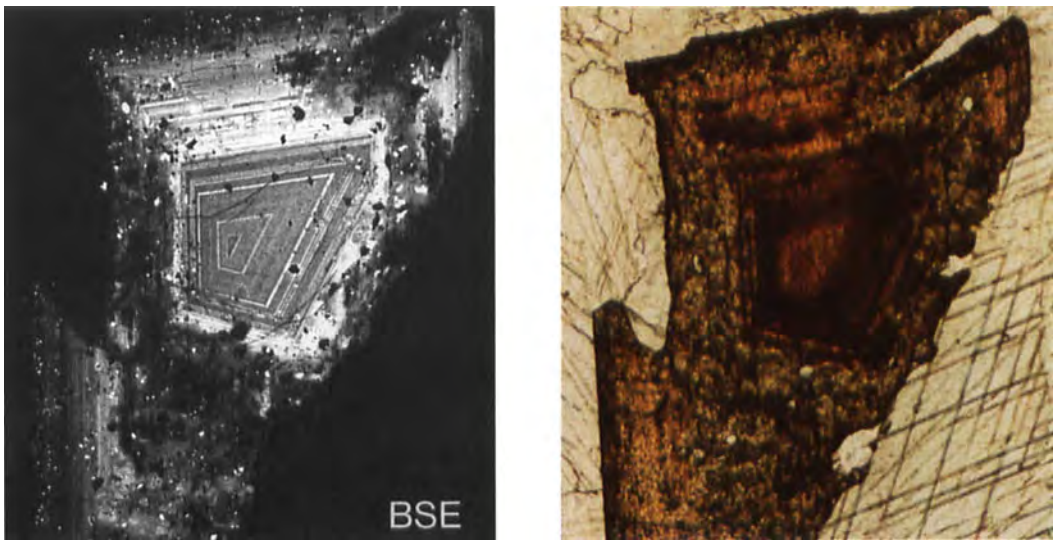


Figure 9. Pyrochlore crystal (30 μm polished section) from magnesiocarbonatite at Fen, Norway (sample from D. D. Hogarth, Ontario): (a) BSE image revealing overall compositional variation obtained with an SEM operating at 15 kV; (b) LOM (transmission) image; X-ray maps for (c) Ce(Lα); (d) Th(Mα); (e) Ta(Lα); and (f) Si(Kα), taken using an electron microprobe operating at 20 kV and 100 nA. Courtesy of Williams, Natural History Museum, London.

methods and ZAF correction; the value of applying the method of principal components analysis to sets of EDS image data is demonstrated elsewhere [77].

Several groups have automated EPMA or SEM analysis for particular mineralogical purposes. WDS detection for precious metals searching was an early development [78]; an improved system with BSE imaging and EDS detection for gold microprospecting is described by Harrowfield et al. [79].

X-ray mapping is useful for revealing major element distributions, compositional zoning, etc., although it lacks sensitivity and is only qualitative. Figures 9a–f show BSE, optical, and X-ray images for Ce, Th, Ta, and Si from a pyrochlore crystal from a carbonatite. The initial oscillatory zoning (Ce varies most strongly) is subsequently partially embayed and corroded. The altered pyrochlore is hydrated, generally richer in Th, and is associated with intergrown silicate phases. The BSE image (a) is a map of overall compositional variation, and being a surface effect, is much sharper than the optical image (b).

2.7 Transmission Electron Microscopy

2.7.1 General Uses

Transmission electron microscopy (TEM) was scarcely applied to geological materials until the 1970s. The use of TEM by relatively few groups has greatly advanced knowledge about natural processes at the grain, subgrain, and atomic scales. Drits [80] and McLaren [81] give good

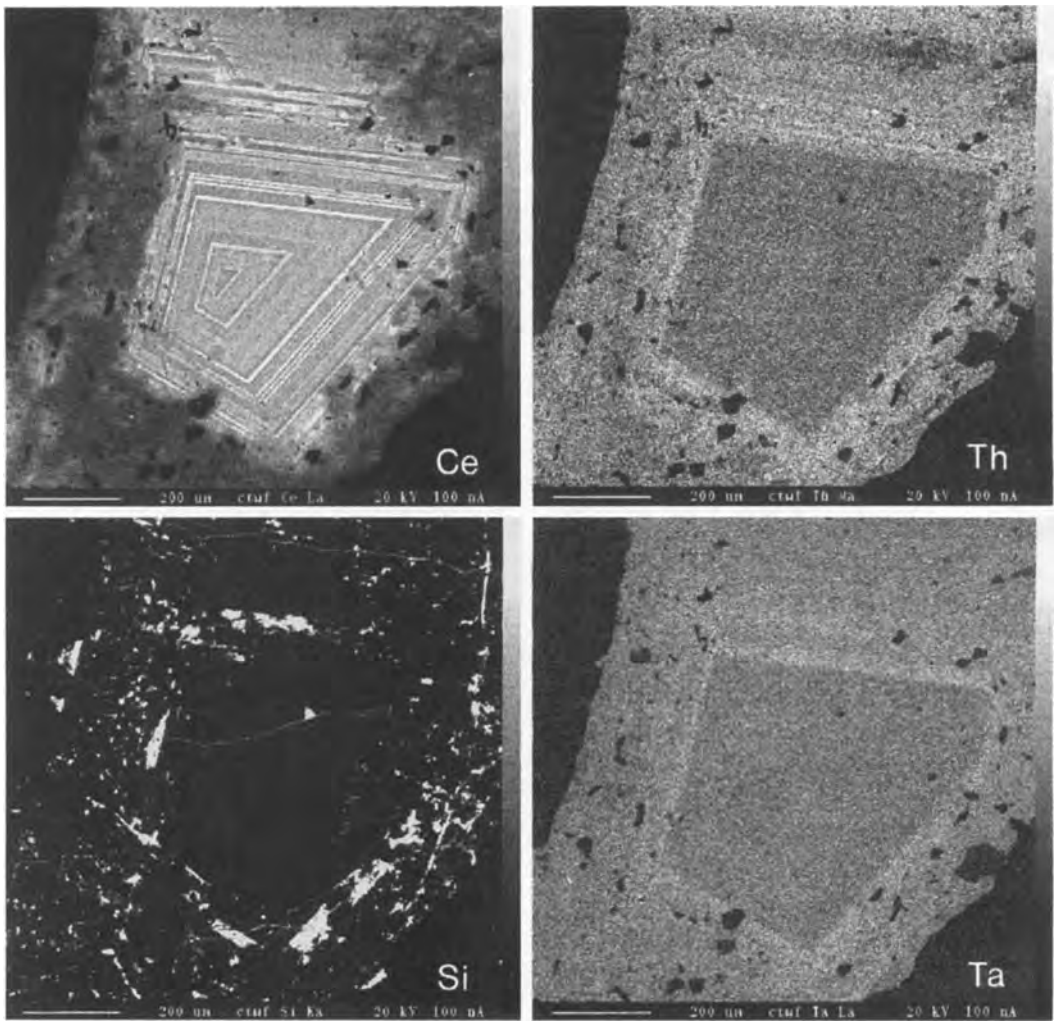


Figure 9. (c)–(f).

introductions to the field; two recent books [82, 83] present very different aspects of the present state-of-the-art.

The advantages of microanalysis in TEM (mainly improved spatial resolution and simpler correction procedures) have been promoted widely. The fine-grained nature of some geological samples makes thin film analysis essential. Currently there are too few examples of the acquisition of high quality TEM/EDS data, although qualitative analyses are common. Applications of ALCHEMI to minerals are discussed in the literature [84, 85]. Automated thin film analysis has been developed [86] for the rapid characterization and the classification of interplanetary dust particles.

Electron energy loss spectrometry (EELS) has not been widely applied to

mineralogical problems, despite its value for light element analysis, and the information it gives about bonding, [85, 87]. EELS data from meteorite residues rich in Cδ, a type of diamond, illustrate the last point [88]; another EELS study of interstellar diamonds from residues was carried out by Bernatowicz et al. [89].

Instrumentation for high resolution electron microscopy (HREM) is not yet sufficient to resolve anions of most oxides and many minerals. This should be possible with the next generation of instruments, where the resolution is expected to be 0.1 nm. Meanwhile, it is possible to determine structures at below the instrumental limit by using three-dimensional image information. This has been demonstrated with staurolite, a resolution of 0.16 nm being achieved [90].

2.7.2 Exsolution, Microstructures, Transformations, and Atomic Structures

Feldspars have played a pivotal role in the development of ideas about differentiation. An abundant and complex group with a wide range of compositions, the feldspars define well the stages of evolution of igneous rocks: crystallization, subsolidus transformation, and alteration. Brown and Parsons [91, 92] have variously addressed exsolution textures, coarsening kinetics, alteration reactions, and twinning in alkali, plagioclase and ternary feldspars. The results have greatly increased knowledge about feldspar microstructures, the mechanisms by which they form, and their relevance. Figure 10a, a micrograph of a cryptomesoperthite from an alkali feldspar rim in a zoned ternary feldspar is an example of the many beautiful TEM images deriving from the work. ALCHEMI and convergent beam electron diffraction (CBED) have been used to study symmetry and Al-Si order, respectively, for a tweed microstructure in an orthoclase (potassium–feldspar) [93]. Since the preparation of TEM specimens is always time-consuming, more rapid methods for revealing fine microstructures in feldspars are often needed. Figures 10b and c illustrate an etching method [94] which enables the scale and type of exsolution to be visualized by SEM. Figure 10b, a TEM image of a similar area from the same specimen, is included for comparison.

Microstructures in calcite and dolomite have been characterized by TEM [95]. Growth and alteration microstructures are considered in [96]. Notable are the elucidation of the nature of ribbon-like microstructures in impure calcites [97] and HREM imaging of various superstructures in calcian dolomite [98], ankerite [99], magnesian calcite [100], and in terms of proposed models [98]. Figure 11a shows the complexity of recent carbonate sediments revealed by HREM, while Fig. 11b shows comparison of simulated images with observed structures of host and a superstructure domain for a calcium ankerite.

TEM has greatly added to knowledge about phase transformations that derived from X-ray diffraction studies, and particularly on the evolution of microstructures with time and temperature. It has demonstrated the common occurrence of metastable, disordered phases outside their stability fields. Phase transformations and transformation-induced microstructures

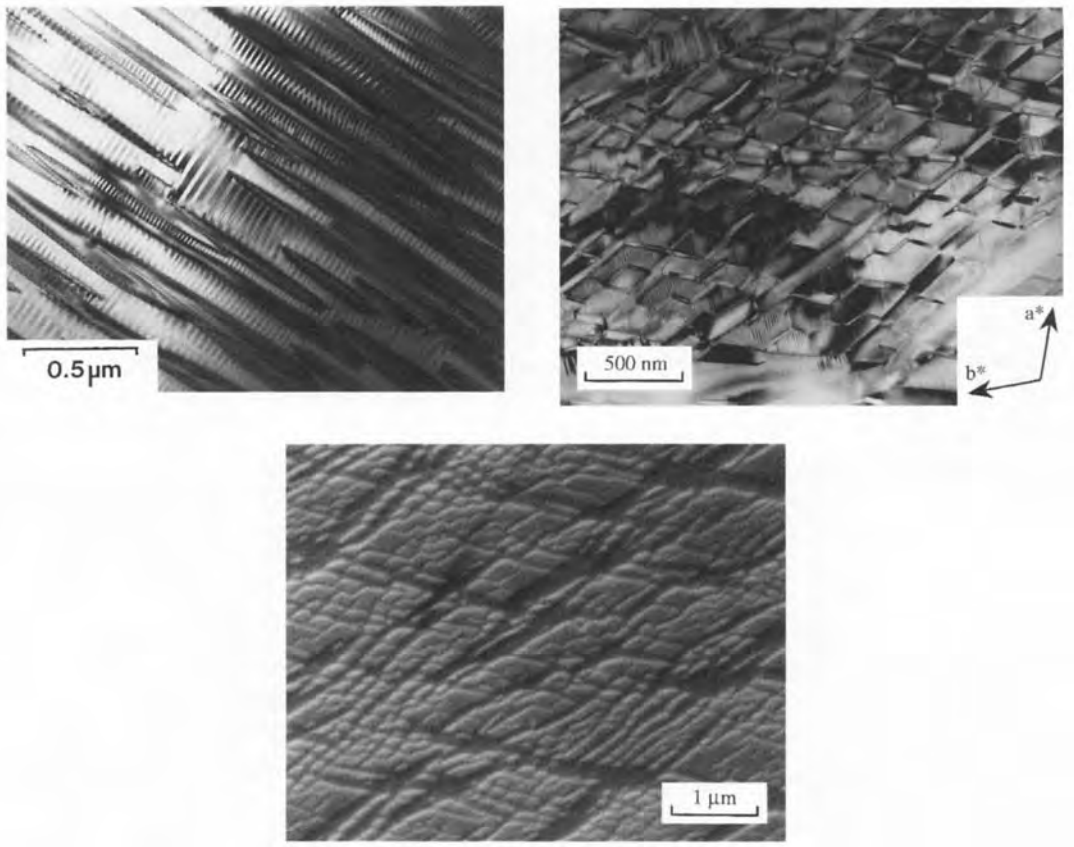


Figure 10. (a) TEM bright field image of cryptomesoperthite from alkali feldspar rims from zoned ternary feldspars. This image shows an orthoclase-poor area where low sanidine occurs in lenses parallel to $(\bar{6}\,0\,1)$. The periodicity of the Albite twinning (lamellae orthogonal to the darker lenses) varies at the ends of the lenses. Courtesy of Brown and Parsons [91]. (b) TEM bright field image and (c) SEM secondary electron image of coherent braid microperthitic alkali feldspar in which zig zag lamellae of low microcline outline lozenges of Albite-twinned low albite. In (b) the electron beam is approximately parallel to the *c*-axis. In (c), the (0 0 1) cleavage surface has been etched with hydrofluoric acid for ≈50 s, which has attacked the low albite lozenges more rapidly, so that they show in relief. Courtesy of Waldron, Lee and Parsons [93]. All three feldspar samples are from the Klokken layered syenite intrusion, South Greenland. Reproduced by permission of Springer Verlag.

in minerals are reviewed by Nord [101]; no summary is attempted here. In situ TEM studies (see also Chap. 2, Sec. 2 of this Volume) of phase transformations in minerals are quite rare. Those of the α–β quartz transformation [102, 103] and of the $P\bar{1}$–$I\bar{1}$ phase transition in anorthite [104] are notable. The first found an incommensurate phase when very close to the transition temperature T_c and produced striking images, as shown in Fig. 11c. The results on the $P\bar{1}$–$I\bar{1}$ transition led to models of the dynamics [105]. An investigation with different aims has shown that in quartzite rocks the Dauphiné twins formed at T_c interact strongly with dislocations and thus can be retained to low temperatures [106].

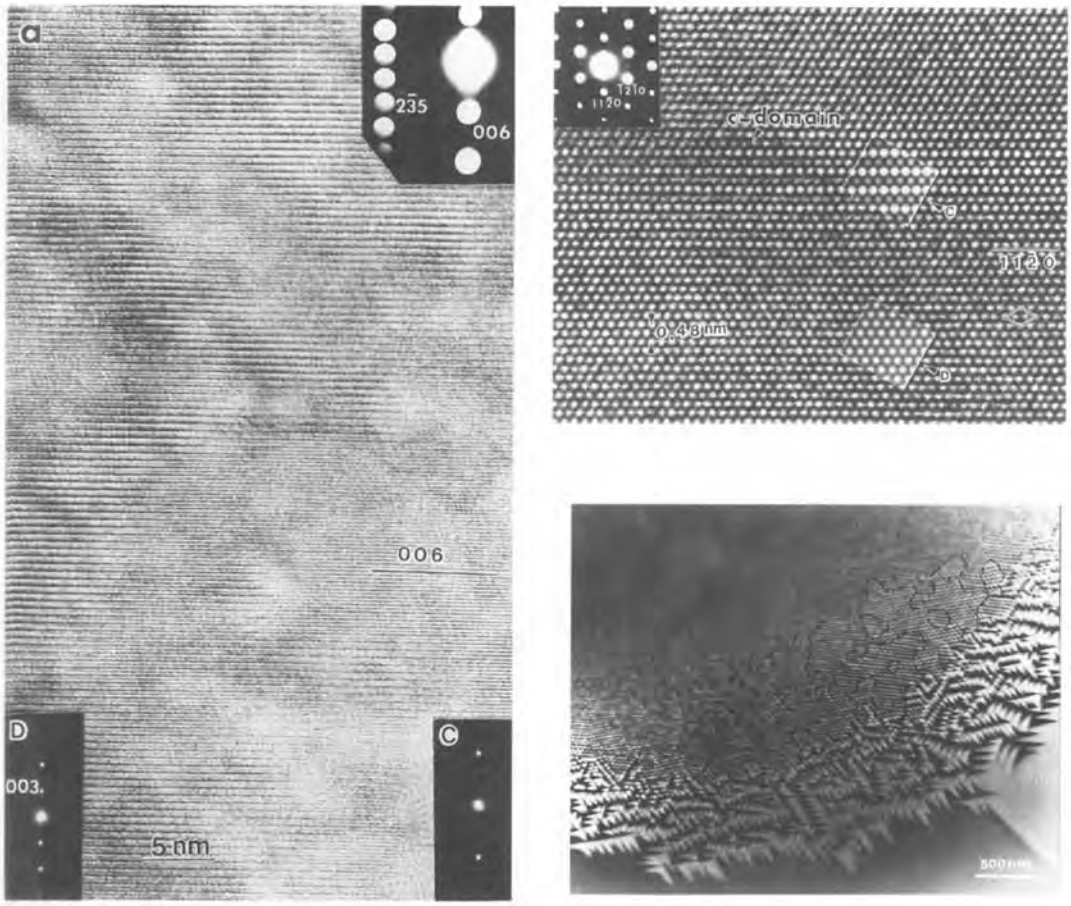


Figure 11. (a) HREM image of Recent Ca–Mg carbonate sediment from Abu Dhabi Sabkha. Spherical domains (within host Mg-calcite) showing fringes characteristic of dolomite ordering. Inset optical diffraction patterns (at bottom corners) from the image show order reflections for dolomite domains (D), but only fundamental reflections for host Mg-calcite (C). Courtesy of Wenk [100]. (b) Comparison of calculated multibeam image simulations with experimental HREM [0 0 0 1] zone images of calcian ankerite and **c** superstructure within it. Simulation at **c** corresponds to the superstructure model proposed in [98]. The simulation at **d** is for ideal dolomite; a hexagonal unit cell is shown for reference. Courtesy of Wenk [99]. (c) Dark field TEM image of the transition region from α-phase to the incommensurate phase close to the phase transition in a quartz specimen cut parallel to the basal plane; (3 0 $\bar{3}$ 1) reflection. The α-phase lies beyond the top left corner. The dark and light triangles are Dauphiné-twin-related prism-shaped domains, and the strong contrast is due to difference in the modulus of the structure factor for α_1 and α_2. Note the differently oriented macrodomains of the incommensurate phase. Courtesy of Van Landuyt [102]. Figures (a) and (b) reproduced by permission of Springer Verlag.

2.7.3 Phyllosilicates

Electron microscopy enables clay minerals to be characterized in three ways: grain texture (grainsize, morphology, by TEM), crystal structure (including stacking sequences, by HREM and diffraction), and chemical composition (by AEM). Most clay minerals present difficulties because they tend to lose bound water

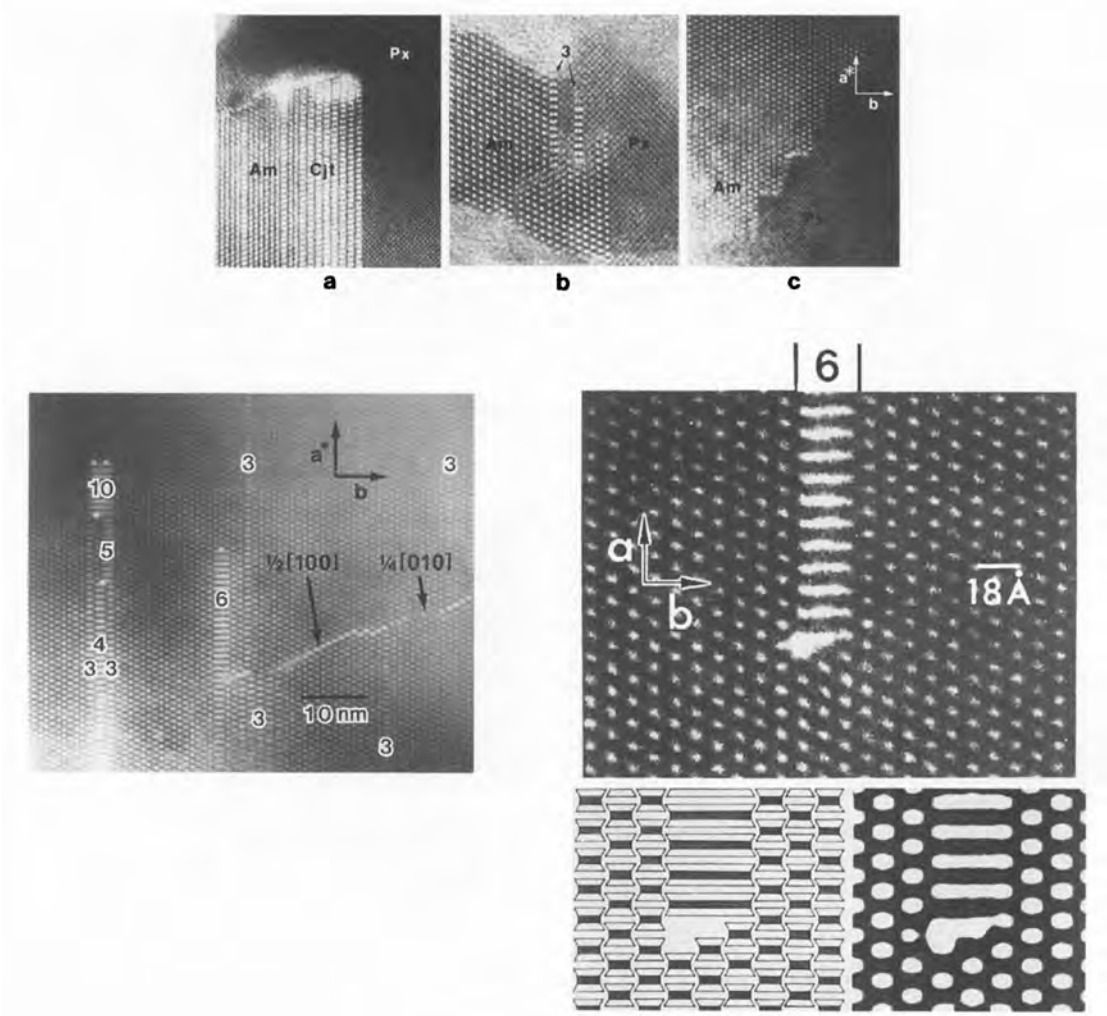


Figure 12. TEM images of replacement reactions in Palisades augite: (a) intergrown amphibole and triple-chain silicate (clinojimthompsonite–Cjt) replacing pyroxene near a fracture; (b) amphibole and augite together with only limited amounts of triple-chain structure (3); (c) the boundary between an amphibole bleb and host augite. The two pyriboles share common crystallographic axes, but the boundary is irregular; (d) part of an amphibole bleb showing wide-chain zippers and displacive faults typical of amphiboles undergoing reactions to sheet silicates. (e) A sextuple-chain defect terminating in the amphibole anthophyllite. TEM image, Jeol JEM 100B. Bottom: structural image in (left) I-beam representation, and (right) computer-simulated image for the same microscope. (a)–(d) Courtesy of Veblen [115]; (e) Courtesy of Veblen [113, 116].

under vacuum and when heated; they are also susceptible to damage by electron and ion irradiation [107]. Their preparation usually requires special techniques and some care.

Disaggregation and dispersal of a typical phyllosilicate onto carbon-coated grids produces platy or fibrous crystals with a small spread of orientations. To view the basal planes 'edge-on' is important, since

this provides greater distinguishing characteristics than when basal planes lie in the plane of the samples. Specimens are therefore prepared either by microtomy or by ion-milling of properly oriented materials. There is much excellent TEM work on phyllosilicates: on crystal structures and variants [108, 109], mixer layer structures [110], hydrothermal alteration and weathering [111], transformations [112], solid state reactions [113], and polytypism [114]. Veblen's work on polysomatism is especially notable. He has shown that reactions in which the reactant and the product belong to the same polysomatic series are common, with pyroxene being replaced by oriented amphibole, wide chain pyriboles, and sheet silicates during alteration and weathering [113]. Figures 12a–c illustrate the replacement of augite by amphibole and triple-chain silicate [115]. Alteration often appears to take place by the growth of lamellae of different chain width ('zippers') into the host. HREM combined with image simulation, Figs. 13a–d of reactions in the double-chain silicate anthophyllite, reveals the natures of the various displacement defects that can terminate zippers when rules governing replacement chain widths are not obeyed [116].

The identification of different phyllosilicate phases and layer sequences can be problematic because of closely similar *d*-spacings, for example, smectites and chlorites, collapsed smectite layers and illites. The utilization of overfocused lattice images and the simulation of images of various phases, Guthrie and Veblen [117, 118], has been influential. The approach has been successful with highly ordered interstratified illite/smectites, and with other mixed layer sequences [107]. Cryotransfer and HREM observation at liquid nitrogen temperatures to prevent collapse of layer spacings is one possible alternative [119]. Chemical treatments have also been developed to stabilize layers against collapse under intense irradiation and otherwise to assist in phase identification. Alkylammonium cations intercalated with phyllosilicate layers arrange themselves to satisfy the layer charge. Another benefit is greater resistance to electron beam heating. HREM measurements of the interlayer spacings of various alkylammonium-saturated phyllosilicates enable the interlayer charge to be estimated and the phases to be identified. This approach is valuable with extraterrestrial phyllosilicates [120], where complex mixtures and interlayering are common, and amounts of material are very small (see this Volume, Chap. 1, Sec. 8).

2.7.4 Deformation Mechanisms in Minerals and Rocks

Microscopy plays a pivotal role in the elucidation of rock deformation mechanisms and the processes giving rise to the observed macro- and micro-structures. The value of LOM for the analysis of petrofabrics, slip mechanisms, and twinning was established many years ago, and it is still valuable for the analysis of deformation in situ [121]. Now both SEM and TEM methods are frequently employed, but optical microscopy is a necessary complement. The scope of such studies ranges from the determination of deformation mechanisms using single crystals, through experimental deformation of rock samples, to the collection and analysis of field

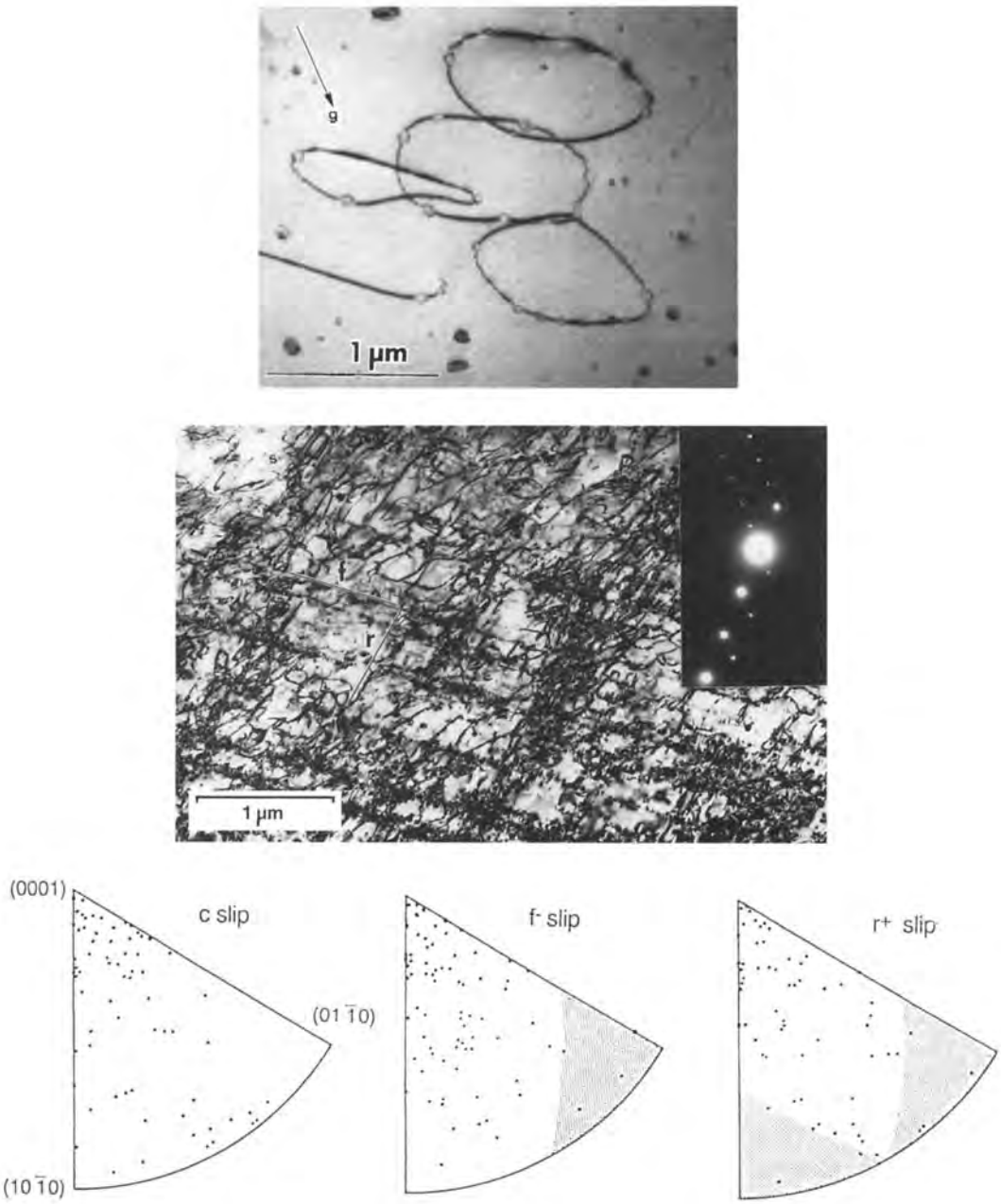


Figure 13. (a) Bright field transmission electron image showing bubbles on dislocation loops in a specimen of quartz heated at 600 °C for 90 min. Diffraction vector, $g = \bar{1}\,1\,0\,1$, the scale mark indicates 1 μm. Courtesy of McLaren [129]. (b) TEM image showing slip on the $(\bar{1}\,0\,1\,2) = f$ and $(1\,0\,\bar{1}\,4) = r$ planes in polycrystalline dolomite deformed in compression at 700 °C under confining pressure; $1\,0\,\bar{1}\,4$ systematic reflection. The slip systems operating and the distribution of deformation were analyzed in some 200 grains from samples deformed at different temperatures; (c) inverse pole figures (equal area projection) combining high and low temperature data representing the final orientations of compression axes for grains deforming by the labeled slip systems; shaded Schmid factor for the expected senses of $(\bar{1}\,0\,1\,2)$ and $(1\,0\,\bar{1}\,4)$ slip, respectively. Barber [132]. Figure (a) reproduced by permission of Springer Verlag and figures (b) and (c) reproduced by permission of Elsevier Science S.A.

samples leading to the interpretation of tectonic events.

Examples of the determination of slip mechanisms using TEM are papers on feldspars [122, 123], quartz [124], diopside [125], dolomite [126], etc. TEM has been very effective in studies of the experimental deformation of polycrystalline aggregates and rocks [127]. Green [128] reviews TEM results on deformation mechanisms in silicates and some high pressure silicate analogs.

Quartz and quartzites deserve special mention, for reasons of history and because TEM has illuminated the problem of hydrolytic weakening, that is, the low strength of 'wet' quartz. Experiments on quartz deformation under a confining pressure led to the earliest use of TEM to the study of mineral deformation. The discovery of hydrolytic weakening further stimulated studies of microstructural effects, often using synthetic quartz to give more control over variables. The results [129–131] from which Fig. 13a derives, are testament to the power of electron microscopy in solving problems in mineralogy.

Assumptions about deformation mechanisms and plasticity which are used in simulations of texture development have been the subject of TEM investigation using deformed dolomite [132]. Figures 13b and c are from this work, which enabled the construction pole figures of the distributions of final orientations of compression axes for grains yielding by particular mechanisms. The number of slip mechanisms active in each grain was always less than the number assumed for simulations and observed subgrain-scale heterogeneity of deformation is not built into current models.

TEM has not been applied widely to field samples of deformed rocks. Mylonitic rocks have received the most attention [133–135]. The work of Knipe [136] is proof that TEM, in combination with optical microscopy, can be a key to the understanding of the deformation within larger scale systems.

2.8 Electron Microscopy of Extraterrestrial Minerals and Rocks

Meteorites and more exotic particles like interplanetary dust particles (IDPs) are our only directly accessible evidence for solar system processes that have made life possible. Although such matter is abundant in space, only a small proportion of that which encounters the Earth's atmosphere is recovered. Consequently, starting with the samples returned by the Apollo missions, techniques requiring only small sample volumes have been pushed to their limits. Maximum information must be extracted from extraterrestrial materials because (i) they result from chance events, so that a given sample may be unique, and (ii) the most informative techniques (ion imaging, TEM) are destructive. Not surprisingly, cutting-edge microscopy is applied to tiny amounts of precious and rare extraterrestrial matter.

SEM/EDS/WDS and EPMA are used routinely in the investigation of the complex mineralogies of stony meteorites, enabling petrologic classifications based on optical microscopy to be extended and refined. Microanalysis has identified meteorites attributed to Martian and to

lunar origins. The differing chemistries of chondrules revealed by EPMA are now attributed mainly to differences in the composition of the solar nebula in various chondrule-forming regions [137]. Thus, electron probe results, together with isotopic data obtained by ion probes, now allow properties of the nebula to be inferred. The enigmatic Ca-Al rich refractory inclusions (CAIs) within chondrules in CV3 carbonaceous chondrites, generally argued to be very primitive [138, 139], have been another fertile field for state-of-the-art EPMA and SEM. Their microstructural complexities and extreme variations in microchemistry would remain undiscovered but for the spatial resolution of present-day electron probes. There are many exemplary papers on this topic, an obvious choice being the first systematic study of rim sequences on CAIs [140]. Later work confirmed their complexity, providing composition profiles, elucidating the reactions responsible [141].

TEM has proved particularly valuable in unraveling the very complex mineralogies of carbonaceous chondrites of low petrologic grade. These contain abundant phyllosilicates, sulfides, and some relict grains. The matrices of meteorites are generally too fine grained to be analyzed by LOM and EPMA methods, and for many years their natures and origins remained obscure. TEM has been crucial in several ways, for example, it shows that much of the layer-silicate matrix of CM chondrites is Fe-rich serpentine. SEM and TEM characterization together indicate that CM and CO matrices, although primitive, have undergone aqueous alteration in an asteroidal regolith [142].

Complex intergrown materials in the CM chondrites, revealed by LOM and SEM, were for some years called PCP (poorly characterized phases). Microanalysis narrowed their identity by showing PCP to be rich in iron, sulfur, silicon and oxygen. A major component was eventually identified as tochilinite [143], a dispersed terrestrial mineral among the first to have its structure determined by electron diffraction [144]. The presence of nierite (Si_3N_4) in four chondrites was discovered by TEM [145], while trace minerals and unsuspected phases in meteorite matrices, for example, saponite, has been identified by using HREM [146]. Small amounts of high melting point silicates in CAIs have been very elusive. One of these, a Ti-rich silicate, has been identified by combining several analytical methods, with convergent beam electron diffraction playing a decisive role [147].

Tiny grains of diamond, other forms of carbon, silicon carbide, and refractory carbides which formed before the solar system and which were incorporated, unmelted, into meteorites have been extracted and identified recently by TEM methods. Their presence was indicated by isotopic anomalies in residues from dissolved meteorites some twenty years ago, but their isolation and identification has proved difficult. The results [88, 148, 149] on such 'stardust' (dust from carbon-rich circumstellar clouds) are both unusual and of great value to the astrophysics community. Figure 14 is an image of the central section through interstellar graphite spherule extracted from the Murchison carbonaceous chondrite. The central crystal is TiC, which acted as a heterogeneous nucleation site for graphite condensation. The sequence TiC $\rightarrow$ graphite is predicted for equilibrium thermodynamics

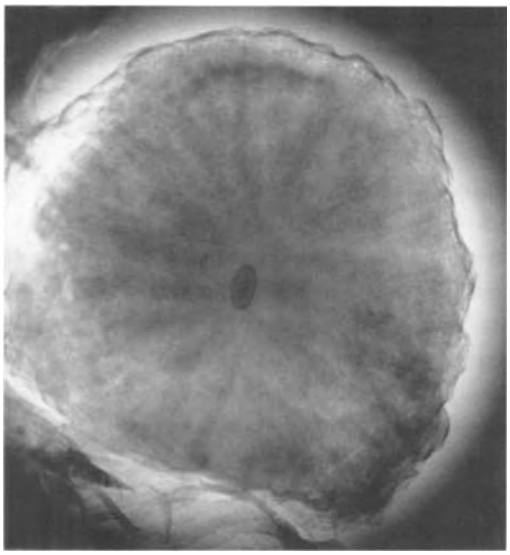


Figure 14. TEM image of an ultrathin section (≈100 nm thick) through an interstellar graphite spherule (≈0.7 μm dia.) extracted from the Murchison meteorite. The central crystal is TiC, which was a nucleation site for graphite condensation. Courtesy of Bernatowicz [149].

for stellar atmospheres with $C/O \approx 1.1$ [148]. TEM has also played a major role in the characterization of interplanetary dust particles [85, 150]. Magneli phases (Ti_nO_{2n-1}) and TiC have been identified within extraterrestrial particles separated from ancient Antarctic ice [151].

2.9 Electron Microscopy of Biogenic Minerals

Biominerals are very different from their geological counterparts in form and properties. The mechanisms of formation and of biological control of microstructure are complex and still somewhat obscure [152]. The microscopy of biominerals is a large subject with some outstanding technical achievements, and it cannot be covered comprehensively here. Much of the optical microscopy and SEM studies can be considered as standard methodology, but it would be a serious omission to mention some of the TEM-based research.

Much attention has naturally been given to hydroxyapatite (HA), the major mineral constituent of bone and teeth. For bone, dentin and mineralized tendon, TEM is essential since the crystals are tiny, typically with lengths and widths ≈10 nm and thicknesses ≈1 nm [153]. Dental enamel has larger HA crystals. Their assemblages and the process of demineralization have been widely studied by TEM and lattice imaging [154].

The mechanisms whereby biominerals form and their microstructures are controlled have also attracted much attention. SEM and TEM have laid the foundations for the present state of knowledge [150]. TEM has been used to analyze the structural complexities [155] and crossed lamellar structures [156, 157] which gives the high strength for weight of molluscan shells, partially inspired by the goal of designing new architectures in synthetic materials based on natural microstructures in biominerals (biomimetics). X-ray microanalysis does not feature much, since it cannot measure the nonstoichiometry and subtle compositional effects that undoubtedly contain many clues, but lattice imaging is used increasingly: for example, to show the single domain nature of magnetosomes of magnetite in magnetotactic bacteria [158] and that barium sulfate crystals in *desmids* are single crystals [159]. HREM plays a growing role in the study of interfaces [160] and in elucidating

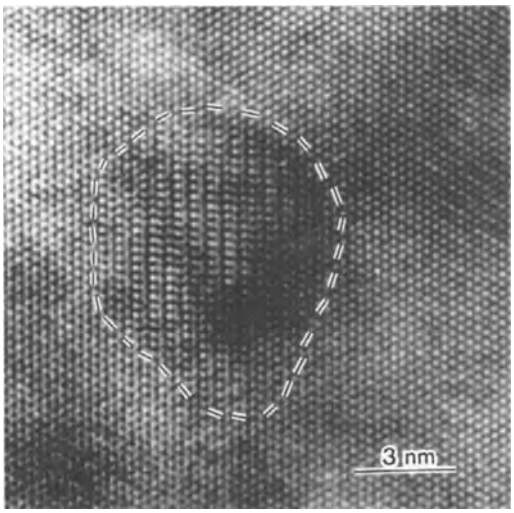


Figure 15. A 400 kV [0 0 *l*] HREM image from calcite from the spine of the deep water echinoderm *Stronglocentrotus franciscanus*. The outlined region is a domain that displays an ordered superstructure and is believed to originate during growth. Courtesy of Buseck [161].

subtle microstructural effects, for example, mosaicism and partially-ordered domains in the calcitic plates and spines of sea urchins [161]. Figure 15 shows a superstructure domain in calcite from the spine of a deep water *echinoderm* [161]. Such are apparently related to mosaicism, but the mechanisms for their formation are not yet understood.

2.10 Auger Electron Spectroscopy (AES) and X-Ray Photoelectron Spectroscopy (XPS)

AES and XPS have not been as widely applied to geological problems as in materials science generally. This may be attributed to a combination of lack of accessibility and exposure of earth scientists to the techniques, to less concern with surfaces as opposed to bulk, and to some technical difficulties (e.g., charging of mineral specimens with AES). Surface analysis has been applied to four main topics in the earth sciences: (i) the oxidation states of near-surface atoms; (ii) adsorption on mineral surfaces; (iii) alteration and weathering; (iv) atomic structures. Hochella [162] summarizes the techniques and reviews results.

The use of XPS and AES in mapping or 'imaging' modes are relatively new. The optimum spatial resolution of a dedicated scanning Auger microprobe, $\approx$70 nm, cannot be achieved with most minerals because they become charged. Effective charge neutralization of the scanning beam is difficult. The application of a *very* thin layer of a material with a low atomic number (e.g., carbon) is one approach, but it usually masks much of the desired information, since the surface specificity of AES is very high. Nonetheless, useful data on complex processes can be obtained by Auger microscopy, as illustrated by a study of the dissolution of labradorite [163].

XPS is generally favored over AES for minerals and XPS more easily provides information about the chemical state of the surface. However, imaging with XPS has much lower resolution than AES, since with XPS either the specimen or the detector must be scanned. A typical spatial resolution for an acceptable signal-to-noise ratio is 70 μm. This might be a tolerable resolution for studies of types (i) and (ii) above, but in general it is limiting.

2.11 Ion Microprobe Analysis and Ion Microscopy

The ion microprobe (dynamic secondary-ion mass spectrometry (SIMS)) is widely used by geologists. It can be employed to measure isotope ratios on very small volumes, and, like EPMA, to define grains of different compositions [164]. Both exploit the ion probe's excellent elemental sensitivity and good spatial resolution. An O^- primary beam is frequently used because rock samples are insulating.

Dynamic SIMS enables the sites of isotopes of particular significance to be identified and ion microscopy enables the spatial distribution of isotopes to be imaged. The discrimination of the sites and concentrations of different isotopes by secondary ion imaging (SII) is very important for investigations of processes which produce isotopic fractionation. Natural processes ranging from terrestrial weathering to stellar nucleosynthesis produce characteristic isotopic ratios. SII is also useful in laboratory studies on minerals where an isotopic tracer is employed, for example, ^{18}O can be a tracer in hydrothermal experiments and has been used to map microporosity and to study micropermeability in feldspars [165]. Ion imaging of ^{18}O after mineral-fluid exchange experiments allows quantification of the relative importance of solute-reprecipitation against solid-state migration and has further benefits in diffusion studies [166].

Equally revealing and state-of-the-art are SII results from the analysis of extraterrestrial materials. The ion probe has been instrumental in identifying mineral grains within primitive meteorites that have anomalous isotopic enrichments (e.g., of ^{26}Mg and ^{17}O [167]), or evidence of extinct radioactivities [168]. Other grains recognized by their isotopic signatures are various types of circumstellar grains that have survived incorporation into the solar system, for example, interstellar SiC has large enrichments of C and N and a nonsolar Si isotopic ratio. Figures 16a,b and c,d are ion images of aluminum oxide and silicon carbide grains, respectively, in residues from meteorite samples dissolved in acids. The oxide grains which show an enrichment of ^{17}O and a depleted ^{18}O (with respect to normal solar system abundances) are deduced to be interstellar grains. Similar arguments apply to silicon carbide grains that have high $^{28}Si/^{30}Si$ ratios [169].

2.12 The Scanning Proton Microprobe and Micro-Proton-Induced X-Ray Emission Spectroscopy

Proton-induced X-ray emission spectroscopy (PIXE) is recommended as a trace element tool, with a better sensitivity than EPMA by a factor of 10 or more [170]. The scanning proton microprobe facilitates X-ray emission analysis at a similar spatial resolution to EPMA. Micro-PIXE (μ-PIXE), as this is known, was pioneered for geological samples at Heidelberg. μ-PIXE has been used by the Heidelberg group both in isolation and to supplement EPMA data [170]. As with EPMA, there is a requirement for homogeneous mineralogical standards which can be a problem [171] in the absence of reliable synthetic ones.

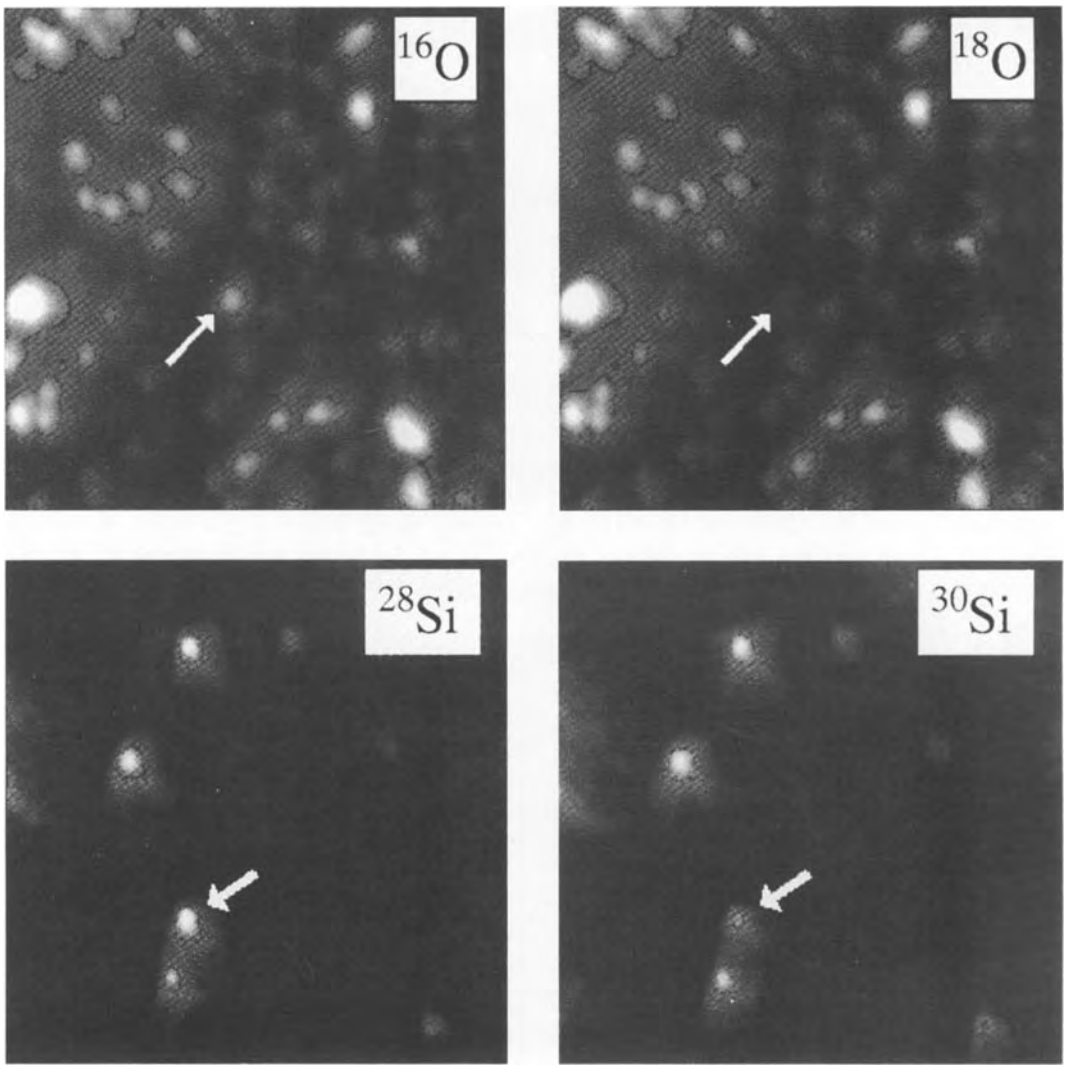


Figure 16. Secondary ion images of dust grains separated from acid residues of primitive meteorites: (a)(b) oxygen isotopes in 1–5 μm Al_2O_3 grains; (c)(d) silicon isotopes in 3 μm SiC grains. Exposure times are selected so that equal intensities reflect normal isotopic compositions. Such ion images allow the rapid identification of rare interstellar dust grains which survived the formation of the solar system, trapped in meteorites. The grains indicated by arrows have isotopic ratios completely different from materials of solar system origin, and they probably formed in the cool atmospheres around giant red stars. Courtesy of Nittler [169].

In addition to the application of μ-PIXE to fluid inclusions [32] (see also Sec. 2.2.3.2 of this Chapter) it has been used for the geothermometry of Ni-bearing garnets, used in assessing diamond-bearing formations [172]. Other geological uses of μ-PIXE are summarized in [173].

2.13 Scanning Tunneling Microscopy and Atomic Force Microscopy

Scanning tunneling microscopy (STM) is of limited use with geological materials because most are too insulating for a tunneling current to flow. STM has therefore mainly been applied to sulfides [174]. The cell of the (1 0 0) galena (PbS) cleavage plane is square with the same dimensions as the bulk unit cell [175]; Pb and S sites are distinguishable, with preferential imaging of the S sites [174]. Defects attributed to Ag and Sb impurity ions suggest that the images can be truly atomic [176]. The (0 0 1) surface cell of hematite is hexagonal, and corresponds closely to the bulk unit cell dimensions [175]. Other minerals imaged by STM include pyrite, stannite and magnetite. The observed atomic structures of the surfaces are in close agreement with those of the corresponding bulk minerals. STM is possible with clay minerals provided that they are ultrathin ($\approx$1 nm) [177].

STM images represent distributions of electron density and should be interpreted in conjunction with calculations about such distributions. STM gives little indication of the composition of imaged features when surfaces are topographically complex. Electron tunneling spectroscopy [178] gives additional information provided that the STM tip is stable (e.g., results on galena [179]).

Atomic force microscopy (AFM) is more versatile and it is increasingly used for the study of mineral surfaces, especially their dynamic processes. The first AFM images of insulators were of boron nitride [180]. AFM studies have to date mostly imaged the surfaces of minerals that cleave easily: graphite, molybdenum disulfide, other sulfide minerals, mica, and calcite. Albite, chlorite, hematite, gibbsite, hydroxyapatite, quartz, montmorillonite, and mixed-layer smectite/illite minerals have also been imaged. The results indicate that surface reconstruction, common with metals and semiconductors, is not prevalent in minerals. This conclusion may be misleading because the surfaces of minerals have been studied in air or under water, and not in ultraclean environments. Also AFM images may not be truly atomic in resolution, for example, the earliest images of calcite may show 'averaged' structures, since no defects were imaged. However, images at 'nearly' atomic resolution have shown a surface cell which is in agreement with both the bulk unit cell and the results from LEED [181]. Figure 17, showing the structure of a calcite cleavage surface, is

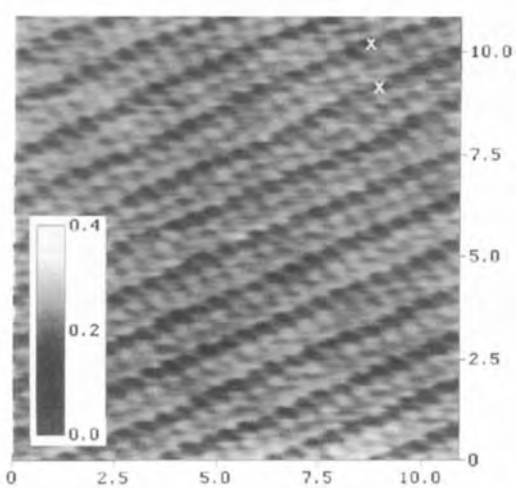


Figure 17. An AFM image of the calcite cleavage plane at atomic resolution showing surficial rows of oxygen atoms (light spots) of the carbonate group, with parallel dark rows (×) formed by the subsurface presence of calcium ions. The vertical inset in this figure provides a scale of surface roughness in nm. Courtesy of Xhie, Hong Kong University of Science and Technology.

believed to have such resolution.

AFM has been used to investigate the applicability of crystal growth theories to quartz and calcite. Sequential measurements have been made on quartz surfaces [182]. In situ AFM observations of surface dynamics have been made with fluid flowing over calcite cleavage surfaces. Surface diffusion is not rapid and step-site nucleation is thought to be the growth-controlling factor [183].

The effects of high energy particle irradiation on various materials have also been investigated by STM and AFM. AFM images the entry points of tracks of Kr ions on mica as depressions in the surface with disordered cores [184]. The application of STM and AFM techniques in the earth sciences is in its infancy. Their use and the breadth of applications will increase as suitable instruments become more widely available.

Acknowledgements

In addition to thanking colleagues who supplied me with illustrations, I am pleased to acknowledge essential help from C. M. O'D. Alexander, W. L. Brown, R. S. Clarke, Jr, R. Hutchison, M. R. Lee, A. C. McLaren, K. M. Moulding, W. R. Nieveen, L. R. Nittler, I. Parsons, R. J. Reeder, D. W. G. Sears, C. J. Spiers, C. T. Williams, Jie Xhie, M. E. Zolensky, and the editors.

2.14 References

[1] F. C. Hawthorne (Ed), *Spectroscopic Methods in Mineralogy and Geology, Rev. in Mineralogy 18*, Min. Soc. Amer., Washington DC **1988**.

[2] L. M. Coyne, S. W. S. McKeever, D. F. Blake (Eds), *Spectroscopic Characterization of Minerals and Their Surfaces*, American Chemical Society, Washington DC **1990**.

[3] N. H. Hartshorne, A. Stuart, *Crystals and the Polarizing Microscope*, Edward Arnold, London **1970**.

[4] W. D. Nesse, *Introduction to Optical Mineralogy*, 2nd ed., Oxford University Press, Oxford **1991**.

[5] G. Harwood in *Techniques in Sedimentology* (Ed: M. E. Tucker), Blackwell Scientific Publications, Oxford **1988**, p. 108.

[6] B. Bousfield, *Surface Preparation and Microscopy of Materials*, Wiley, Chichester **1992**.

[7] W. S. Mackenzie, C. Guilford, *Atlas of Rock-forming Minerals in Thin Section*, Longman, Harlow **1980**.

[8] B. W. D. Yardley, W. S. Mackenzie, C. Guilford, *Atlas of Metamorphic Rocks and Their Textures*, Longman, Harlow **1990**.

[9] R. H. Beauchamp, J. F. Williford in *Metallographic Specimen Preparation: Optical and Electron Microscopy* (Eds: J. A. McCall, W. M. Mueller), Plenum Press, New York **1974**, p. 233.

[10] K. Fredriksson, A. F. Noonan, J. Nelen, R. Beauchamp, *Meteoritics* **1974**, *13*, 462.

[11] D. J. Barber, *Mineral. Mag.* **1981**, *44*, 357.

[12] F. J. Turner, *Nat. Acad. Sci. Proc.* **1962**, *48*, 955.

[13] F. J. Turner, L. E. Weiss, *Structural Analysis of Metamorphic Tectonites*, McGraw-Hill, New York **1963**.

[14] W. D. Carlson, J. L. Rosenberg, *J. Geol.* **1981**, *89*, 615.

[15] G. P. Price, *J. Geol.* **1980**, *88*, 181.

[16] X. Bai, P. Smart, X. Leng, *Géotechnique* **1994**, *44*, 175.

[17] P. Ramdohr, *The Ore Minerals and Their Intergrowths*, 2nd. ed., Pergamon Press, Oxford **1980**.

[18] R. E. Reed-Hill, in *Techniques of Metals Research* (Ed: R. F. Bunshah), Vol. 2, Wiley Interscience, New York **1968**, p. 257.

[19] T. Lu, I. Sunagawa, *Phys. Chem. Minerals* **1990**, *17*, 207.

[20] I Sunagawa, T. Lu, V. S. Balitsky, *Phys. Chem. Minerals* **1990**, *17*, 320.

[21] H. J. Milledge, G. P. Bulanova, W. R. Taylor, P. A. Woods, P. H. Turner, *Abstracts: 6th Int. Kimberlite Conf., Novosibirsk, Russia* **1995**, p. 384.

[22] H.-J. Schubnel, M. Pinet, D. C. Smith, B. Lasnier, *La Microsonde Raman en Gemmologie*, Association Française de Gemmologie, Paris **1992**.

[23] R. J. Hemley, P. M. Bell, E. C. T. Chao, *Phys. Chem. Minerals* **1986**, *13*, 285.

[24] S. Y. Wang, S. K. Sharma, T. F. Cooney, *Am. Mineral.* **1993**, *78*, 469.

[25] E. Roedder (Ed), *Fluid Inclusions, Rev. Mineral. 12*, Min. Soc. Amer., Washington DC **1984**.
[26] S. D. Burley, J. Mullis, A. Matter, *Marine Petroleum Geol.* **1989,** *6*, 98.
[27] N. Petford, J. A. Miller, A. Rankin, *J. Microsc.* **1995,** *178*, 37.
[28] P. Dhamelincourt, J. M. Beny, J. Dubessy, B. Poty, *Bull. Mineral.* **1979,** *102*, 600.
[29] G. J. Rosasco, E. Roedder, *Geochim. Cosmochim. Acta* **1979,** *43*, 1907.
[30] J. C. Touray, C. Beny-Bassez, J. Dubessy, N. Guilhaumou, *Scanning Electr. Microsc.* **1985,** 103.
[31] J. D. Pasteris, B. Wopenka, J. C. Seitz, *Geochim. Cosmochim. Acta* **1988,** *52*, 979.
[32] A. J. Anderson, A. H. Clark, X.-P. Ma, G. R. Palmer, J. D. McArthur, E. Roedder, *Economic Geol.* **1989,** *84*, 924.
[33] M. Montoto, A. Martínez-Nistal, A. Rodríguez-Rey, N. Fernández-Merayo, P. Soriano, *J. Microsc.* **1995,** *177*, 138.
[34] N. Petford, J. A. Miller, *Am. Mineral.* **1993,** *77*, 529.
[35] N. Petford, J. A. Miller, J. Briggs, *Comp. Geosciences* **1993,** *19*, 585.
[36] K. Nishiyama, H. Kusuda, *Int. J. Rock Mech. Min. Sci. Geomech. Abstr.* **1994,** *31*, 369.
[37] P.-J. Perie, P. Monteiro, *Int. J. Rock Mech. Min. Sci. Geomech. Abstr.* **1991,** *28*, 83.
[38] P. Holler, F.-C. Kögler, *Marine Geol.* **1990,** *91*, 263.
[39] T. Lu, K. Toyoda, N. Nango, T. Ogawa, *Rev. Laser Eng.* **1991,** *19*, 440.
[40] I. Sunagawa, T. Lu, *Mineral. J. (Japan)* **1987,** *13*, 328.
[41] A. R. Lang, *Diamond Related Mater.* **1993, 2,** 106.
[42] F. C. Frank, J. W. Harris, K. Kaneko, A. R. Lang, *J. Cryst. Growth* **1994,** *143*, 46.
[43] M. Fehlmann, D. Bertelmann in *Synchrotron Radiation Applications in Mineralogy and Petrology* (Ed: A. Barto-Kyriakidis). Theophrastus Publications, Athens **1988,** p. 205.
[44] C. Ilett, M. G. Somekh, G. A. D. Briggs, *Proc. Roy. Soc. Lond.* **1984,** *A393*, 171.
[45] A. Rodriguez-Ray, G. A. D. Briggs, T. A. Field, M. Montoto, *J. Microsc.* **1990,** *160*, 21.
[46] D. Knauss, G. A. D. Briggs, *J. Phys. D.* **1994,** *27*, 1976.
[47] D. J. Marshall, *Cathodoluminescence of Geological Materials*, Unwin Hyman, Boston **1988**.
[48] J. Miller in *Techniques in Sedimentology* (Ed: M. E. Tucker), Blackwell Scientific Publications, Oxford **1988,** Chap. 6, p. 174.
[49] I. D. Hutcheon, I. M. Steele, J. V. Smith, R. N. Clayton, *Proc. Lunar and Planet. Sci. Conf.* IX **1978,** p. 1345.
[50] M. R. Lee, G. M. Harwood, *Sedim. Geol.* **1989,** *65*, 285.
[51] R. J. Reeder in *Luminescence Microscopy and Spectroscopy: Qualitative and Quantitative Applications* (Eds: C. E. Barker, O. C. Kopp), SEPM Short Course 25, SEPM (Soc. Sed. Geol.), Tulsa **1991,** p. 77.
[52] J. Rakovan, R. J. Reeder, *Amer. Mineral*, **1994,** *79*, 892.
[53] I. M. Steele, *Amer. Mineral.* **1986,** *71*, 966.
[54] I. M. Steele, *Geochim. Cosmochim. Acta* **1986,** *50*, 1379.
[55] K. Ramseyer, J. Fischer, A. Matter, P. Eberhardt, J. Gneiss, *J. Sedim. Petrol.* **1989,** *59*, 619.
[56] H. Namba, Y. Murata, *J. Phys. Soc. Japan* **1984,** *53*, 1888.
[57] M. V. Nazarov, T. A. Nazarova, *Microscopy and Analysis* **1995,** *45*, 29.
[58] D. W. G. Sears, J. M. DeHart, F. A. Hasan, G. E. Lofgren in [2], Chap. 11.
[59] K. Ninagawa, S. Nishimura, N. Kubono, I. Yamamoto, M. Kohata, T. Wada, Y. Yamashita, J. Lu, D. W. Sears, S. Matsunami, H. Nishimura, *Proc. NIPR Symp. Antarct. Meteorites* 5 (Ed: K. Yanai), National Institute for Polar Research, Tokyo **1992,** p. 281.
[60] N. H. Trewin in *Techniques in Sedimentology* (Ed: M. E. Tucker), Blackwell Scientific Publications, Oxford **1988,** Chap. 8, p. 229.
[61] W. J. McHardy, M. J. Wilson, J. M. Tait, *Clay Minerals* **1982,** *17*, 23.
[62] P. Smart, N. K. Tovey, *Electron Microscopy of Soils and Sediments: Examples.* Clarendon Press, Oxford **1981**.
[63] P. Smart, N. K. Tovey, *Electron Microscopy of Soils and Sediments: Technique.* Clarendon Press, Oxford **1982**.
[64] W. D. Keller, R. P. Haenni, *Clays Clay Minerals* **1978,** *26*, 384.
[65] P. Tapponnier, W. F. Brace, *Int. J. Rock Mech. Min. Sci. & Geomech. Abstr.* **1976,** *13*, 103.
[66] C. W. Montgomery, W. F. Brace, *Contrib. Mineral. Petrol.* **1975,** *52*, 17.
[67] C. J. Spiers, P. M. T. M. Schutjens in *Deformation Processes in Minerals, Ceramics and Rocks* (Eds: D. J. Barber, P. G. Meredith), Unwin Hyman, London **1990,** Chap. 12.
[68] G. E. Lloyd, R. D. Law, S. M. Schmid, *J. Struct. Geol.* **1987,** *9*, 251.
[69] G. E. Lloyd, N. Ø. Olesen, N.-H. Schmidt, *Scanning* **1988,** *10*, 163.
[70] N. Ø. Olesen, N.-I. Schmidt in *Deformation Mechanisms, Rheology and Tectonics* (Eds: R. J. Knipe, E. H. Rutter), Spec. Publ. No. 54, Geol. Soc. London **1990,** p. 369.
[71] K. Z. Baba-Kishi, D. J. Dingley, *Scanning* **1989,** *11*, 305.

[72] D. J. Dingley, K. Z. Baba-Kishi, V. A. Randall, *Atlas of Electron Backscatter Kikuchi Patterns*, IOP Publishing, Bristol **1995**.
[73] K. Kunze, B. L. Adams, F. Heidelbach, H.-R. Wenk, *Proc. Icotom 10, Mater. Sci. Forum* **1994,** *157–162*, 1243.
[74] K. Kunze, F. Heidelbach, H.-R. Wenk, B. L. Adams in *Textures of Geological Materials* (Eds: H. J. Bunge, S. Siegesmund, W. Skrotski, K. Weber), Deutsche Ges. Metallk., Oberursel **1994,** p. 127.
[75] A. E. Bence, A. L. Albee, *J. Geol.* **1968,** *76*, 382.
[76] J. T. Armstrong in *Microbeam Analysis – 1988* (Ed: D. E. Newbury), San Francisco Press, San Francisco, **1988,** p. 469.
[77] J. M. Paque, R. Browning, P. L. King, P. Pianetta, Proc. 12th International Congress on Electron Microscopy (Eds: L. D. Peachey, D. B. Williams), Vol. 2, San Francisco Press, San Francisco **1990,** p. 244.
[78] M. D. McConnell, R. B. Bolon, W. T. Grubb in *Microbeam Analysis – 1983* (Ed: K. F. Heinrich) San Francisco Press, San Francisco **1983,** p. 125.
[79] I. R. Harrowfield, C. M. MacRae, P. F. Simmonds in *Microbeam Analysis – 1988* (Ed: D. Newbury), San Francisco Press, San Francisco **1988,** p. 481.
[80] V. A. Drits, *Electron Diffraction and High Resolution Electron Microscopy of Mineral Structures*, Springer Verlag, Berlin **1981**.
[81] A. C. McLaren, *Transmission Electron Microscopy of Minerals and Rocks*. Cambridge University Press, Cambridge **1991**.
[82] J. N. Boland, J. D. Fitz Gerald (Eds), *Defects and Processes in the Solid State: Geoscience Applications – The McLaren Volume*, Elsevier, Amsterdam **1993**.
[83] P. R. Buseck (Ed), *Minerals and Reactions at the Atomic Scale, Rev. Mineral. 27*, Min. Soc. America, Washington DC **1992**.
[84] I. D. Mackinnon in L. M. Coyne, S. W. S. McKeever, D. F. Blake (Eds), *Spectroscopic Characterization of Minerals and Their Surfaces*, American Chemical Society, Washington DC **1990,** Chap. 2.
[85] P. R. Buseck, P. Self in *Minerals and Reactions at the Atomic Scale*, (Ed: P. R. Buseck), *Rev. Mineral. 27*, Min. Sec. America, Washington DC **1992,** Chap. 5.
[86] J. P. Bradley, M. S. Germani, D. E. Brownlee, *Earth Planet. Sci. Lett.* **1989,** *93*, 1.
[87] K. M. Krishnan in [2], Chap. 3.
[88] D. F. Blake, F. Freund, K. F. M. Krishnan, C. J. Echer, R. Shipp, T. E. Bunch, A. G. Tielens, R. J. Lipari, C. J. D. Hetherington, S. Chang, *Nature* **1988,** *332*, 611.
[89] T. J. Bernatowicz, P. C. Gibbons, R. S. Lewis, *Astrophys. J.* **1990,** *359*, 246.
[90] H. R. Wenk, K. H. Downing, Hu Meishing, M. A. O'Keefe, *Acta Cryst.* **1992,** *A8*, 700.
[91] W. L. Brown, I. Parsons, *Contrib. Mineral. Petrol.* **1988,** *98*, 444.
[92] W. L. Brown, I. Parsons in *Feldspars and Their Reactions* (Ed: I. Parsons), Kluwer, Amsterdam **1994,** p. 449.
[93] K. Waldron, M. R. Lee, I. Parsons, *Contrib. Mineral. Petrol.* **1994,** *116*, 360.
[94] A. C. McLaren, J. D. Fitz Gerald, *Phys. Chem. Minerals* **1987,** *14*, 281.
[95] H.-R. Wenk, D. J. Barber, R. J. Reeder in *Carbonates: Mineralogy and Chemistry* (Ed: R. J. Reeder), *Rev. Mineral. 11*, Min. Soc. America, Washington DC **1983,** Chapter 9.
[96] R. J. Reeder, in [83], Chap. 10.
[97] D. J. Barber, M. R. Khan, *Mineral. Mag.* **1987,** *51*, 71.
[98] G. Van Tendeloo, H.-R. Wenk, R. Gronsky, *Phys. Chem. Minerals* **1985,** *12*, 333.
[99] H.-R. Wenk, H. Meisheng, T. Lindsey, J. W. Morris, *Phys. Chem. Minerals* **1991,** *17*, 527.
[100] H.-R. Wenk, H. Meisheng, S. Frisia, *Am. Mineral.* **1993,** *78*, 769.
[101] G. L. Nord, in [83], Chap. 12.
[102] G. Van Tendeloo, J. Van Landuyt, S. Amelinckx, *Physica Status Solidi (a)* **1976,** *33*, 723.
[103] N. Yamamoto, K. Tsuda, K. Yagi, *J. Phys. Soc. Jpn.* **1988,** *57*, 2559.
[104] S. Ghose, G. Van Tendeloo, S. Amelinckz, *Science* **1988,** *242*, 1539.
[105] G. Van Tendeloo, S. Ghose, S. Amelinckx, *Phys. Chem. Minerals* **1989,** *16*, 311.
[106] D. J. Barber, H.-R. Wenk, *Phys. Chem. Minerals* **1991,** *17*, 492.
[107] D. R. Peacor in [83], Chap. 9.
[108] D. R. Veblen, P. R. Buseck, C. W. Burnham, *Science* **1977,** *198*, 359.
[109] S. Guggenheim, R. A. Eggleton in *Hydrous Phyllosilicates* (Ed: S. W. Bailey), *Rev. Mineral. 19*, Min. Soc. America, Washington DC **1988,** p. 675.
[110] J. H. Ahn, D. R. Peacor, *Clays Clay Minerals* **1989,** *37*, 542.
[111] J. F. Banfield, R. A. Eggleton, *Clays Clay Minerals* **1988,** *36*, 47.
[112] M. Amouric, S. Parc, D. Nahon, *Clays Clay Minerals* **1991,** *39*, 254.
[113] D. R. Veblen in [83], Chap. 6.
[114] A. Baronnet in [83], Chap. 7.
[115] D. R. Veblen, P. R. Buseck, *Am. Mineral.* **1981,** *66*, 1107.
[116] D. R. Veblen, P. R. Buseck, *Am. Mineral.* **1980,** *65*, 599.

[117] G. D. Guthrie, D. R. Veblen, *Clay Minerals* **1989,** *37*, 1.
[118] G. D. Guthrie, D. R. Veblen in [2], Chap. 4.
[119] T. Murakami, T. Sato, T. Watanabe, *Am. Mineral.* **1993,** *78*, 465.
[120] D. W. Ming, S. V. Yang, D. C. Golden, K. A. Thomas, L. P. Keller, K. Krivian, R. A. Barrett, M. E. Zolensky, *Proc. Lunar Planet. Sci. Conf. XXII* **1992,** p. 913.
[121] L. Urai, *Tectonophysics* **1987,** *135*, 251.
[122] D. B. Marshall, A. C. McLaren, *Phys. Chem. Minerals* **1977,** *1*, 351.
[123] M. Gandais, C. Willaime in *Feldspars and Feldspathoids* (Ed: W. L. Brown), D. Reidel, Dordrecht **1984,** p. 207.
[124] J.-C. Doukhan, C. Trepied, *Bull. Mineral.* **1985,** *108*, 97.
[125] P. Raterron, N. Doukhan, O. Jaoul, J.-C. Doukhan, *Phys. Earth Planet. Interiors* **1994,** *82*, 209.
[126] D. J. Barber, H. C. Heard, H.-R. Wenk, *Phys. Chem. Minerals* **1981,** *7*, 271.
[127] J. Tullis in *Deformation Processes in Minerals, Ceramics and Rocks* (Eds: D. J. Barber, P. G. Meredith), Unwin Hyman, London **1990,** Chap. 8.
[128] H. W. Green in [83], Chap. 11.
[129] A. C. McLaren, R. F. Cook, S. T. Hyde, R. C. Tobin, *Phys. Chem. Minerals* **1983,** *9*, 1979.
[130] J. D. Fitz Gerald, J. N. Boland, A. C. McLaren, A. Ord, B. E. Hobbs, *J. Geophys. Res.* **1991,** *96*, 2139.
[131] J. Gerretsen, A. C. McLaren, M. S. Paterson in [82], p. 27.
[132] D. J. Barber, H.-R. Wenk, H. C. Heard, *Mat. Sci. Eng.* **1994,** *A175*, 83.
[133] S. H. White, D. J. Evans, D.-L. Zhong, *Textures Microstruct.* **1982,** *5*, 33.
[134] S. J. Hippler, R. J. Knipe in *Deformation Mechanisms, Rheology and Tectonics* (Eds: R. J. Knipe, E. H. Rutter), Special Publication 54, Geol. Soc. London **1990,** p. 71.
[135] L. B. Goodwin, H.-R. Wenk, *Tectonophysics* **1990,** *178*, 389.
[136] R. J. Knipe in *Deformation Processes in Minerals, Ceramics and Rocks* (Eds: D. J. Barber, P. G. Meredith), Unwin Hyman, London **1990,** Chap. 9.
[137] J. N. Grossman, A. E. Rubin, H. Nagahara, E. A. King in *Meteorites and the Early Solar System* (Eds: J. F. Kerridge, M. S. Matthews), University of Arizona Press, Tucson, **1988,** p. 619.
[138] G. J. MacPherson, D. A. Wark, J. T. Armstrong in *Meteorites and the Early Solar System* (Eds: J. F. Kerridge, M. S. Matthews), University of Arizona Press, Tucson, **1988,** p. 746.
[139] A. Hishimoto, L. Grossman, *Geochim. Cosmochim. Acta* **1987,** *51*, 1685.
[140] D. A. Wark, J. F. Lovering, *Proc. Lunar. Sci. Conf. VIII*, **1977,** p. 95.
[141] X. Hua, J. Adam, H. Palme, A. El Goresy, *Geochim. Cosmochim. Acta* **1988,** *52*, 1389.
[142] E. R. D. Scott, D. J. Barber, C. M. Alexander, R. Hutchison, J. A. Peck in *Meteorites and the Early Solar System* (Eds: J. F. Kerridge, M. S. Matthews), Univ. of Arizona Press, Tucson, **1988,** p. 718.
[143] I. D. R. Mackinnon, M. E. Zolensky, *Nature* **1984,** *309*, 240.
[144] N. I. Organova, V. A. Drits, A. L. Dmitrik, *Kristallografiya* **1972,** *17*, 761.
[145] M. R. Lee, S. S. Russell, J. W. Arden, C. T. Pillinger, *Meteoritics* **1995,** *30*, 387.
[146] K. Tomeoka, P. R. Buseck, *Geochim. Cosmochim. Acta* **1990,** *54*, 1745.
[147] D. J. Barber, J. R. Beckett, J. M. Paque, E. Stolper, *Meteoritics* **1994,** *29*, 682.
[148] P. Fraundorf, G. Fraundorf, T. Bernatowicz, R. Lewis, M. Tang, *Ultramicroscopy* **1989,** *27*, 401.
[149] T. J. Bernatowicz, S. Amari, E. K. Zinner, R. S. Lewis, *Astrophys. J.* **1991,** *373*, L73.
[150] J. P. Bradley, D. E. Brownlee, *Science* **1991,** *251*, 549.
[151] M. E. Zolensky, A. Pun, K. L. Thomas, *Proc. Lunar & Planet. Sci. Conf. XIX* (Eds: G. Ryder, V. L. Sharpton), Lunar Planet. Inst., Houston **1989,** p. 505.
[152] L. Addadi, S. Weiner, *Angew. Chem. Int. Ed. Engl.* **1992,** *31*, 153.
[153] S. Weiner, P. Price, *Calcif. Tissue Int.* **1986,** *39*, 365.
[154] S. Wen, *Electron Microsc. Rev.* **1989,** *2*, 1.
[155] M. Sarikaya, I. A. Aksay in *Results and Problems in Cell Differentiation 19: Biopolymers* (Ed: St. T. Case), Springer Verlag, Berlin **1992,** Chap. 1.
[156] V. J. Laraia, M. Aindow, A. H. Heuer in *Mater. Res. Symp. Proc. 174* (Eds: P. C. Rieke, P. Calvert, M. Alper), MRS, Pittsburgh **1990,** p. 117.
[157] N. Wilmot, D. J. Barber, J. Taylor, A. L. Graham, *Philos. Trans. R. Soc. Lond.* **1992,** *B337*, 21.
[158] F. C. Meldrum, S. Mann, B. R. Heywood, R. B. Frankel, D. A. Bazylinski, *Proc. R. Soc. Lond.* **1993,** *B251*, 237.
[159] J. R. Wilcock, C. C. Perry, R. J. P. Williams, A. J. Brook, *Proc. R. Soc. Lond.* **1989,** *B238*, 203.
[160] D. J. Barber, *Scripta Metall. Mater.* **1994,** *31*, 989.
[161] S. J. Tsipursky, P. R. Buseck, *Am. Mineral.* **1993,** *78*, 775.

[162] M. F. Hochella in [2], Chap. 13.
[163] M. F. Hochella, H. B. Ponader, A. M. Turner, D. W. Harris, *Geochim. Cosmochim. Acta* **1988,** *52*, 385.
[164] R. Levi-Setti, J. M. Chabala, Y. Wang, P. Hallégot, C. Girod-Hallégot in *Images of Materials* (Eds: D. B. Williams, A. R. Pelton, R. Gronsky), Oxford University Press, Oxford **1991,** p. 94.
[165] F. D. L. Walker, *Contrib. Mineral. Petrol.* **1990,** *106*, 124.
[166] S. C. Elphick, C. M. Graham, F. D. L. Walker, M. B. Holness, *Mineral. Mag.* **1991,** *55*, 347.
[167] I. D. Hutcheon, G. R. Huss, A. J. Fahey, G. J. Wasserburg, *Astrophys. J.* **1994,** *425*, L97.
[168] T. Lee, D. A. Papanastassiou, G. J. Wasserburg, *Geophys. Res. Lett.* **1976,** *3*, 109.
[169] L. R. Nittler, C. M. O'D. Alexander, X. Gao, R. M. Walker, E. K. Zinner, *Nature* **1994,** *370*, 443.
[170] H. Blank, K. Traxel, *Scanning Electr. Microsc. – 1984*, **1984,** 1089.
[171] C. G. Ryan, D. R. Cousens, S. H. Sie, W. L. Griffin, G. F. Suter, *Nucl. Instrum. Methods* **1990,** *B47*, 55.
[172] W. L. Griffin, D. R. Cousens, C. G. Ryan, S. H. Sie, G. F. Suter, *Contrib. Mineral. Petrol.* **1989,** *103*, 199.
[173] J. L. Campbell, W. J. Teesdale, N. M. Halden, *Geoscience Canada* **1992,** *19*, 175.
[174] C. M. Eggleston, M. F. Hochella, *Geochim. Cosmochim. Acta* **1990,** *54*, 1511.
[175] M. F. Hochella, C. M. Eggleston, V. B. Elings, G. A. Parks, G. E. Brown, Jr, C. M. Wu, K. Kjoller, *Am. Mineral.* **1989,** *74*, 1223.
[176] T. G. Sharp, N. J. Zheng, I. S. T. Tsong, P. R. Buseck, *Amer. Mineral.* **1990,** *75*, 1438.
[177] H. Lindgreen, J. Garnæs, P. L. Hansen, F. Besenbacher, E. Lægsgaard, I. Stensgaard, S. A. C. Gould, P. K. Hansma, *Am. Mineral.* **1991,** *76*, 1218.
[178] J. A. Stroscio, R. M. Feenstra, A. P. Fein, *Phys. Rev. Lett.* **1986,** *57*, 2579.
[179] C. M. Eggleston, M. F. Hochella, *Am. Mineral.* **1993,** *78*, 877.
[180] T. R. Albrecht, C. F. Quate, *J. Appl. Phys.* **1987,** *62*, 2599.
[181] A. L. Rachlin, G. S. Henderson, M. C. Goh, *Am. Mineral.* **1992,** *77*, 904.
[182] A. J. Gratz, S. Manne, P. K. Hansma, *Science*, **1990,** *251*, 1343.
[183] A. J. Gratz, P. E. Hillner, P. K. Hansma, *Geochim. Cosmochim. Acta* **1993,** *57*, 491.
[184] F. Thibaudau, J. Cousty, E. Balanzat, S. Bouffard, *Phys. Rev. Lett.* **1991,** *67*, 1582.

Handbook of Microscopy
S. Amelinckx, D. van Dyck,J. van Landuyt , G. van Tendeloo

3 Semiconductors and Semiconductor Devices

3.1 Introduction

The impressive progress in microelectronics is based on silicon technology using MOS (metal oxide semiconductor) transistors. Since 1969, every three years a new generation of MOS technology emerges with four times as many transistors as the previous generation. At present the leading semiconductor fabrication lines are producing 4 Mbit and 16 Mbit dynamic random access memories (DRAMs) and logic integrated circuits (ICs) with critical dimensions of 0.7 and 0.5 μm, respectively. Sixty-four megabit DRAMs enter the production phase, and functional 256 Mbit samples are already available. Research and development is focusing on 1 Gbit/4 Gbit devices with 0.1 μm structures (theoretical and practical limits of gigascale integration are discussed by Meindl [1]). The scientific and industrial community feels convinced that progress in microelectronics will not encounter fundamental roadblocks for the next 15 years [2]. Apart from miniaturization, further progress will need material innovations, such as new metals and dielectrics. CMOS (complementary MOS) technology is dominating increasingly because of advantages in complexity, power consumption, and speed.

Silicon technology comprises the formation of active device areas in the silicon substrate by doping, the fabrication of dielectric layers, and the deposition of metallization layers. Of course, all regions and layers require patterning by photolithography and etching. For an introduction to silicon technology see, for example, Sze [3] and Ghandi [4]. The active device areas in MOS transistors are source and drain regions as well as the gate channel. Dielectric layers are either grown by thermal oxidation, resulting in thin (<10 nm) or thick (~1 μm) SiO_2 layers, or by chemical vapor deposition (CVD) of silicon dioxide or silicon nitride. The gate and metallization layers involve primarily the materials polycrystalline silicon ('polysilicon'), metal silicides, tungsten and aluminum.

With increasing miniaturization the complexity of the process technology also increases, which places higher demands on techniques for inspection, characterization, or analysis. Miniaturization of course calls for microscopic techniques providing high spatial resolution for both imaging and analysis. A large variety of optical, electron beam, and ion beam methods are employed for both materials characterization during technology development as well as for inspection and failure analysis during process optimization and process

control. Apart from microscopic techniques, methods for ultratrace analysis of inorganic and organic impurities in process media and on the wafer surfaces are also required. For example, metal contamination levels as low as 10^{10} atoms cm^{-2} have to be analyzed.

The main emphasis in microscopic techniques will be on the application of electron microscopy because of the excellent spatial resolution down to nanometers for scanning electron microscopy (SEM) and down to the atomic scale for transmission electron microscopy (TEM). Equally important is the possibility of obtaining reliable results within a short time during routine operation. SEM is most widely used for the imaging of device structures, and allows process control by nondestructive inspection within the fabrication line. Cross-sectional samples for SEM can be easily prepared by cleavage or polishing, and provide detailed information for technology assessment. The cross-section of Fig. 1 was prepared by focused ion beam (FIB) etching and shows two perpendicular section planes through a 16 Mbit DRAM. Prior to SEM imaging the section surface was selectively etched to delineate the various layers more clearly. TEM on the other hand allows more detailed materials characterization, as is visualized in the inserts of the cross-section through a 4 Mbit DRAM (Fig. 2): geometrical details, crystal defects, and material reactions at interfaces. TEM cross-sections in particular provide an understanding of material effects occurring in the various technology processes. In addition to imaging, additional analytical techniques such as energy-dispersive X-ray (EDX) spectroscopy are extremely valuable for both SEM and TEM. The various

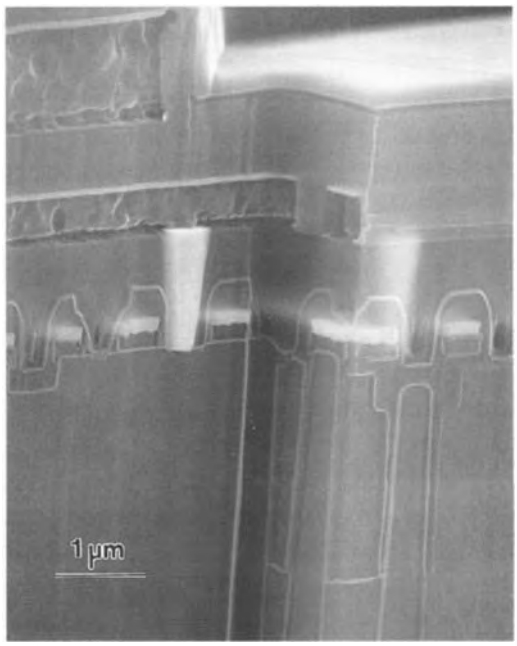


Figure 1. SEM image of a cross-section through a 16 Mbit memory chip (DRAM). Part of a trench capacitor is visible in the lower right corner. The bright 'needle' left of the center is a tungsten plug providing contact between the first aluminum metallization level and the silicon substrate.

scanning probe microscopy techniques, such as STM (scanning tunneling microscopy) and AFM (atomic force microscopy), of course provide superb resolution both laterally and vertically, but are still in an initial stage of routine application. Quantitative determination of surface roughness after cleaning and etching processes and the study of epitaxial growth mechanisms of silicon and compound semiconductors have been major fields of application. In many cases the methods of surface analysis are required. Both Auger electron spectroscopy (AES) and secondary ion mass spectroscopy (SIMS) offer the possibility of spatially resolved chemical analysis. An

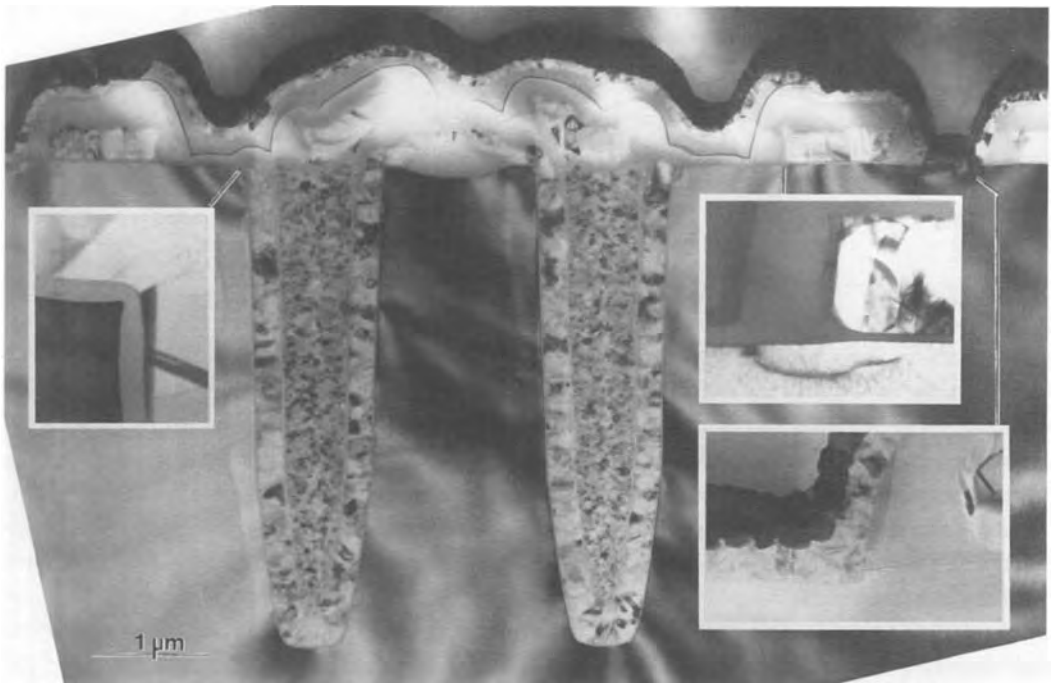


Figure 2. TEM image of a cross-section through the storage cell area in a 4 Mbit DRAM. The insets show magnifications of critical device features which require TEM inspection: oxide thinning at the trench edge (left), dislocations in the silicon substrate, and material reactions in contacts to the substrate (right).

introduction can be found in Fuchs et al. [5]. In failure analysis of electronic devices, various methods for defect localization such as electron beam testing (EBT) and emission microscopy are applied before physical failure analysis by SEM or TEM can be employed to identify the failure cause. Finally, special emphasis must be put on specimen preparation, which is extremely important for TEM. Advanced procedures for preparation of TEM cross-sections of semiconductor samples are described by, for example, Benedict et al. [6]. FIB systems offer the intriguing possibility of both local specimen preparation (e.g., etching of a cross-section, see Fig. 1) and imaging with good resolution. A wealth of detailed information on both microscopic methods and materials science problems may be found in the proceedings of the biannual Oxford conferences on the microscopy of semiconducting materials [7].

As indicated above, the main emphasis is on silicon technology. For compound semiconductors focus is put on characterization of heteroepitaxial layer structures. Such heterostructures are used for both electronic and optoelectronic devices. Since the thickness of the individual layers ranges down to a few nanometers, various methods of cross-sectional TEM imaging are discussed. In addition, examples of GaAs technology are described briefly.

3.2 Semiconductor Bulk Material

3.2.1 Silicon for Device Fabrication

Silicon single crystals for integrated circuit fabrication have been grown dislocation-free by float zone (FZ) and Czochralski (CZ) growth for almost three decades. In quantity, most of the silicon single-crystalline material is grown by the CZ technique, the rods now having a diameter of up to 300 mm. Though FZ and CZ grown silicon crystals are dislocation-free, they contain a variety of grown-in defects. These are agglomerations of self-point defects (silicon interstitials or vacancies) such as dislocation loops and stacking faults. Moreover, CZ silicon contains a very high concentration of oxygen, up to the solid solubility of oxygen in silicon, which arises from the oxygen from the SiO_2 crucible material dissolving into the silicon melt. Thus, CZ silicon contains SiO_x precipitates and their associated secondary defects, that is, dislocation loops and stacking faults. The density and distribution of all these grown-in defects depend strongly on the crystal growth parameters which influence the distribution and concentration of the self-point defects and impurities. In the starting silicon material for device manufacturing, that is, the silicon wafer, grown-in defects have to be avoided in the electrically active device areas.

Characterization of these microdefects involves first imaging of their distribution within the wafer on a larger scale, and second identifying their crystallographic nature in detail. For imaging of defect distributions defect etching using, for example, Secco or Wright–Jenkins etch [8] and optical microscopy is still an extremely sensitive and straight forward method. In addition, techniques based on infrared microscopy have been used successfully. Infrared microscopy at low magnification allows wafer mapping of inhomogeneities. Scanning infrared microscopy is more sensitive and can image individual microdefects much smaller than the resolution by using various imaging modes [9]. An analogous procedure is laser scanning tomography [10]. A narrow beam from an infrared laser is made to enter the edge of a wafer and to run within the wafer parallel to its surface. Light scattered through 90° from the beam is recorded with an infrared camera. Scanning the wafer parallel to the wafer surface in a direction perpendicular to the laser beam yields a two-dimensional (2D) image. Detailed characterization of the defect nature (dislocation, stacking fault, interstitial, or vacancy type) and distribution was originally obtained by X-ray topography of silicon slices cut from the crystal ingot [11]. A more precise analysis of the individual microdefects can be obtained by TEM, and a few examples will be discussed below.

In the periphery of FZ and CZ wafers the silicon self-interstitial concentration is enhanced, which may lead to particular microdefects—the so-called swirls. The A defect type of these swirls were identified by TEM to be interstitial dislocation loops having extensions of up to a few micrometers [12]. By controlling the impurities such as oxygen, carbon, and nitrogen as well as the growth parameters, swirls can be avoided. The demand for thinner gate oxides (<10 nm) makes the oxide more susceptible to degradation by grown-in

defects. Defect etching revealed so-called COPs (crystal originated particles) in the centre of CZ silicon, and FPDs (flow pattern defects) in the center of both FZ and CZ wafers. These defects cause severe problems for gate oxide integrity. TEM analysis revealed the FPDs to be tiny interstitial dislocation loops [13].

In CZ crystal growth, point defect clusters consisting of the impurities oxygen and carbon and the self-point defects are formed, acting as nucleation centers for SiO_x precipitation and stacking fault generation in the cooler parts of the just grown crystal. A homogenization procedure at high temperature dissolves the SiO_x precipitates before silicon wafers are cut from the ingot. In order to avoid oxygen precipitation in the active device area close to the silicon surface, a high-temperature annealing step is applied to diffuse out the oxygen and form an oxygen-denuded zone several micrometers wide below the wafer surface. Then, a low-temperature annealing follows to create a high density of nucleation centers for SiO_x precipitates. Succeeding high-temperature annealing causes the SiO_x precipitates to grow [14]. Extensive TEM analysis was carried out to document the precipitate and secondary defect morphology after the various annealing processes. A full compilation was given in a recent review [15]. The most predominant defects are SiO_x platelets and truncated SiO_x octahedra which are amorphous as revealed by high-resolution electron microscopy (HREM). The composition is substoichiometric SiO_2 as measured with electron energy loss spectroscopy (EELS) [16] and infrared absorption spectroscopy. Since an SiO_2 precipitate requires a 2.25 times larger volume than silicon, every precipitated oxygen atom generates 0.63 silicon interstitials which are released into the silicon bulk. The platelets lie on {1 0 0} planes and have ⟨1 1 0⟩ edges. As long as they are only a few nanometers thick their contrast in a bright-field image is dominated by the strain field (Fig. 3). A 'line of no contrast' appears where the atomic displacements associated with the strain field are perpendicular to the diffraction vector. According to the respective diffraction vectors, the line of no contrast lies along the platelet in Fig. 3a and normal to it in Fig. 3b. Further growth in thickness leads to the generation of punched out interstitial-type prismatic dislocation loops along four different ⟨1 1 0⟩ directions (Fig. 4). In this way the high strain is released by silicon interstitial emission. The characteristic bright-field image of a truncated octahedral precipitate is shown in Fig. 5. The octahedra have {1 0 0} and {1 1 1} faces and are almost strain-free. The silicon interstitials emitted from such a defect can nucleate an extrinsic stacking fault bordered by a Frank partial dislocation. The bright-field image in Fig. 6 shows such a stacking fault having nucleated at an octahedral precipitate. SiO_x precipitates and their secondary defects become harmful to device performance (gate oxide breakdown, and refresh-time killers in DRAMs) when they are generated in a denuded zone that has not been properly produced. The degree of precipitation also strongly determines the wafer warpage. The SiO_x precipitates and their secondary defects in the bulk of the wafer also have a beneficial effect: they serve as gettering sites for heavy metal impurities. In order to optimize this kind of gettering process the mechanism of 'intrinsic' gettering has to be understood. TEM analysis showed

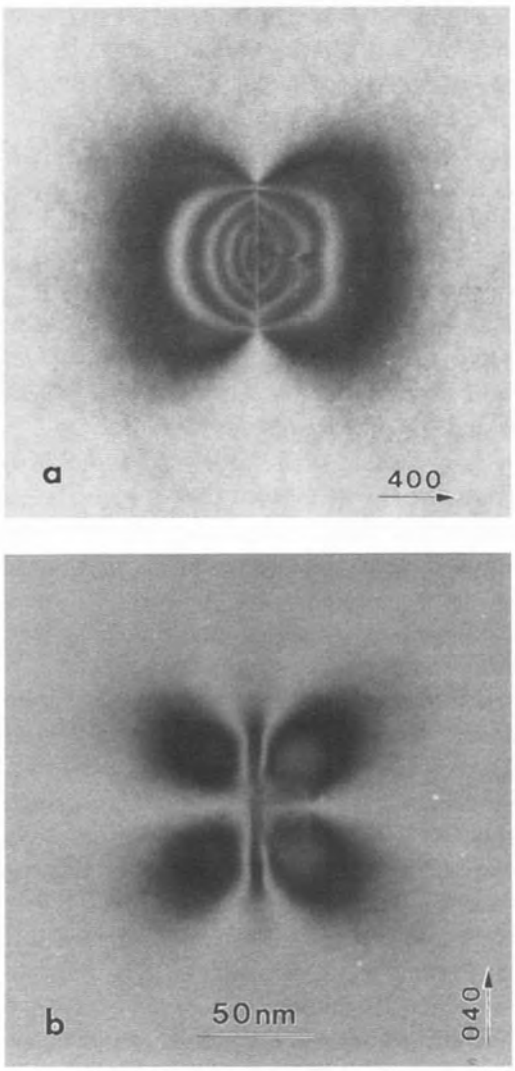


Figure 3. Strain contrast of a thin amorphous SiO_x precipitate in silicon lying on a {1 0 0} plane. Bright field TEM images of the same precipitate recorded with the (4 0 0) (a) and (0 4 0) (b) reflections perpendicular and parallel to the precipitate plane, respectively.

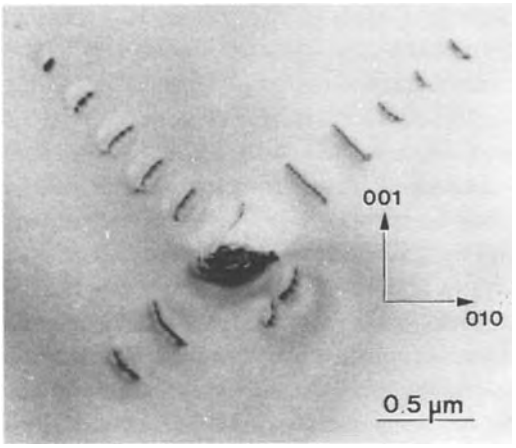


Figure 4. Extrinsic dislocation loops punched out on four equivalent glide systems from a platelike, amorphous SiO_x precipitate in the center of the image.

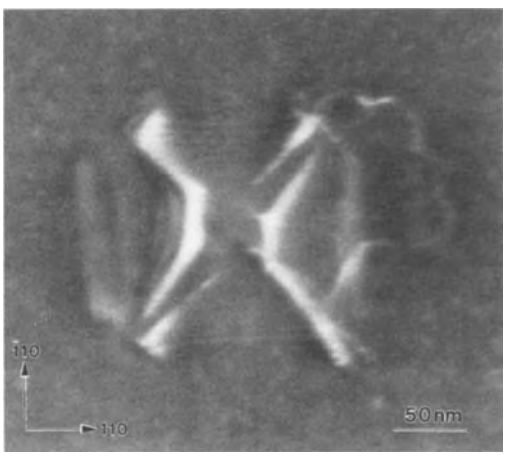


Figure 5. Weak-beam image of an amorphous SiO_x precipitate having the shape of a truncated octahedron with {1 1 1} and {1 0 0} faces.

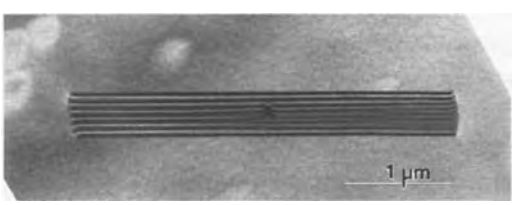


Figure 6. Extrinsic stacking fault accommodating the Si interstitials which were emitted by the SiO_x precipitate at its center.

that heavy metals (iron, nickel, and copper) precipitate as silicides at the SiO_x precipitates [17] or at the prismatic dislocation loops and Frank partial dislocations of the stacking faults [18, 19]. Which

kind of defect acts as a nucleation site depends strongly on the metal silicide, that is, whether it emits or absorbs silicon interstitials on precipitation, and on the strain field of the defect. The homogeneous nucleation of metal silicide precipitates in the bulk of silicon wafers has been studied in detail by Seibt [20–22].

3.2.2 Compound Semiconductors

From the numerous compound semiconductors only a few are used as single-crystal substrate material for semiconductor devices. III–V semiconductors, in particular gallium arsenide (GaAs) and to a lesser extent indium phosphide (InP), find broad application as substrates for both optoelectronic and electronic devices. Optoelectronic devices such as laser diodes and light-emitting diodes exploit the direct bandgap of GaAs and InP, allowing radiative recombination processes of charge carriers leading to light emission. Electronic devices for high frequencies (>10 GHz) such as microwave and millimeter wave power devices benefit from the higher mobility of charge carriers compared to silicon. Devices based on silicon carbide (SiC) hold promise for high-voltage and high-temperature applications but are still at an exploratory stage.

The requirements for GaAs substrate crystals used for optoelectronic devices differ notably from those for electronic devices. Fabrication of optoelectronic devices involves the growth of heteroepitaxial layer structures, the characterization of which will be described in Section 3.5.1 of this Chapter. The substrate crystals have to be highly doped since high currents flow from the back contact through the substrate to the pn junction in the layer structure on top of the wafer. Doping levels are 10^{18}–$10^{19}\,cm^{-3}$ with silicon for n-type and carbon or zinc for p-type GaAs. The dislocation density has to be extremely low since individual dislocations in the active layer of a laser diode enhance nonradiative recombination and the local increase in temperature may degrade the laser. For GaAs crystals grown by the horizontal Bridgman technique, dislocation densities of 1000–$2000\,cm^{-2}$ are usually obtained. Recently, the vertical gradient freeze technique has yielded GaAs wafers with even lower dislocation densities below $500\,cm^{-2}$ [4].

In contrast, for electronic devices, undoped semi-insulating GaAs substrate crystals are required to provide isolation between the active device areas. No useful thermal oxide which allows lateral isolation of active device structures in silicon technology can be grown on GaAs. Undoped, semi-insulating GaAs crystals are predominantly grown by the liquid encapsulated Czochralski (LEC) technique, and contain significant dislocation configurations. The average dislocation density is between 10^4 and $10^5\,cm^{-2}$ and is correlated to the thermal stress operating during crystal growth [23]. The dislocation distribution can be imaged by X-ray topography, infrared microscopy, and defect etching. Figure 7 shows an X-ray topograph of a 75 mm GaAs wafer. Because of the high absorption of the Mo-K_α radiation by the 0.5 mm thick GaAs wafer, only the X-ray wavefield anomalously transmitted through perfect crystal regions causes bright contrast. In regions of high defect density a larger angular width of the incoming X-ray

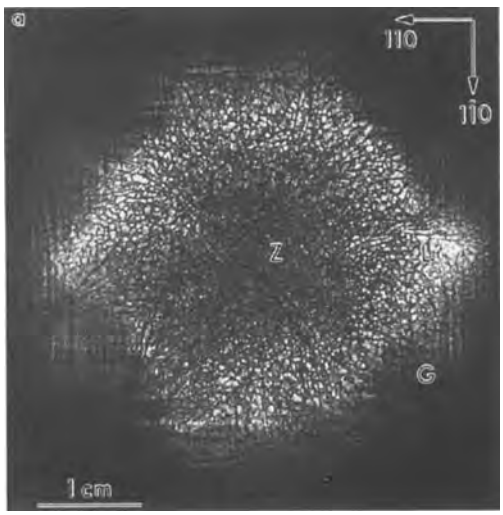


Figure 7. X-ray topograph of an undoped semi-insulating GaAs wafer showing dislocation cell walls in the center (Z), slip bands (G), and lineages (L).

beam is diffracted initially, but is subsequently absorbed in the crystal, and those regions, therefore, appear dark. In the center of the wafer (Z in Fig. 7) the dislocations are concentrated in dislocation cell walls, enclosing virtually dislocation-free material. The cell size is below 100 μm. In the wafer periphery, slip bands parallel to ⟨1 1 0⟩ directions (G in Fig. 7) and lineages develop. In a detailed analysis of defect configurations by TEM [24], it proved advantageous to mark the cell walls or glide bands by chemical etching in order to find individual dislocations by TEM. The DSL etch (diluted Sirtl etch applied with light) which is sensitive to the point defects surrounding the dislocations [25], was used for this purpose. Figure 8 shows a dislocation in a cell wall with two precipitates (P) associated with it. These precipitates exhibit an increased arsenic peak in the EDX spectrum, and were identified to consist of rhombohedral arsenic [26]. In the dark field image of a polycrystalline arsenic precipitate (Fig. 9a) only part of it appears bright. In the corresponding selected-area diffraction pattern (Fig. 9b) no orientation relationship between precipitate and the GaAs matrix can be detected. Analysis of the Burgers vector and the plane of the dislocation in Fig. 8 showed that the dislocation had moved by positive climb, that is, by the addition of vacancies or the emission of interstitials. While arsenic precipitation is assumed to occur soon after solidification at temperatures above 800°C, positive climb of dislocations is proposed to occur at lower temperatures with gallium vacancies providing the driving force for a climb process during which arsenic antisite defects, that is, arsenic atoms on gallium lattice sites, are created [24]. These antisite defects are widely agreed to correspond to the EL2 centers, the main electronic trap in semi-insulating GaAs.

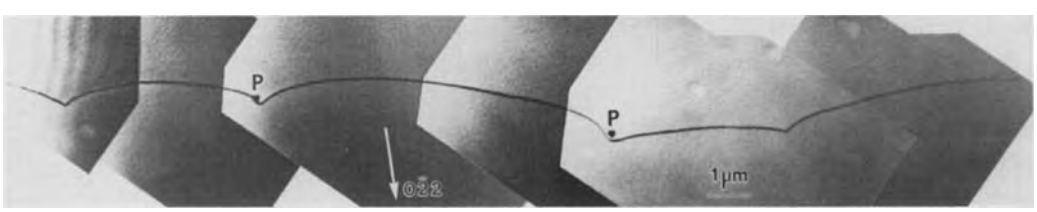


Figure 8. Bright field image of a dislocation in undoped GaAs which was pinned at precipitates (P) during climb.

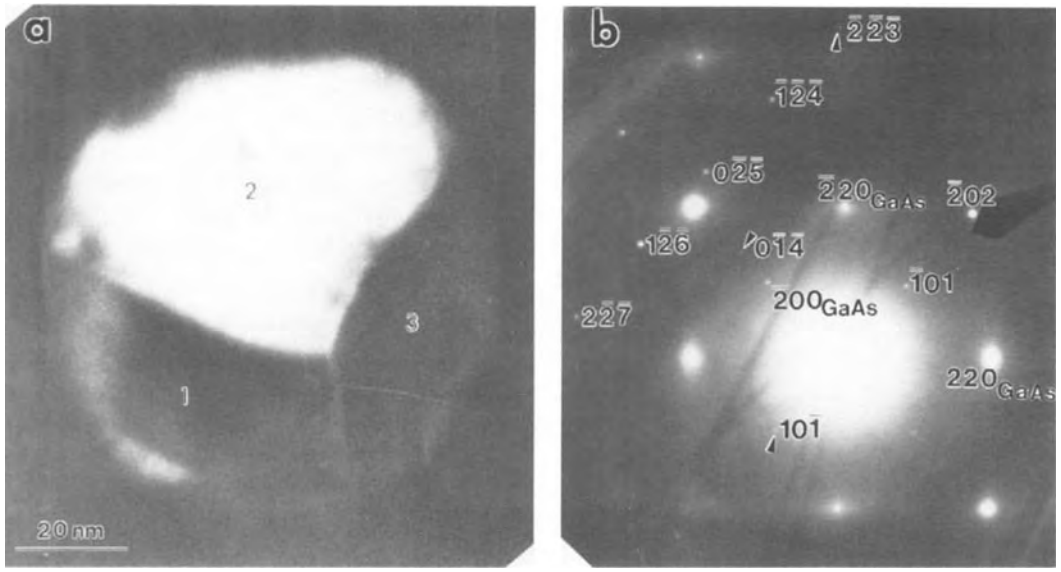


Figure 9. (a) Dark field image of a polycrystalline arsenic precipitate imaged in the $(\bar{2}02)$ reflection. (b) Corresponding selected-area diffraction pattern. Indices without the subscript GaAs refer to rhombohedric arsenic.

Progress in LEC crystal growth of semi-insulating GaAs has led to increased crystal diameters. Up to 100 mm diameter wafers are used in fabrication presently with up to 150 mm diameters in the future. Despite the increase in wafer diameter the dislocation density could be kept to levels of 10^4–10^5 cm^{-2}, which are not critical to performance and reliability of discrete GaAs devices.

3.3 Silicon Technology Processes

3.3.1 Dielectric Layers

Apart from the semiconducting properties of silicon, the possibility of growing high-quality dielectrics by thermal oxidation is a major benefit in silicon technology. The gate oxide of MOS transistors is usually grown by dry oxidation with a thickness of the order of 10 nm or even below. Thin dielectrics are also required for the capacitors in DRAMs. For lateral insulation between different device structures, thick oxides are grown by wet oxidation up to thicknesses around 1 μm (field oxides). In addition to thermal oxidation, chemical vapor deposition is used to fabricate silicon dioxide and silicon nitride layers. Such dielectrics are used for insulation between metallization layers and for passivation to protect devices from impurities, moisture, and scratches.

Since the interface between the gate oxide and the silicon substrate has an influence on device properties, for example via the interface trap density and the interface roughness, the structure and composition of this interface has been studied in

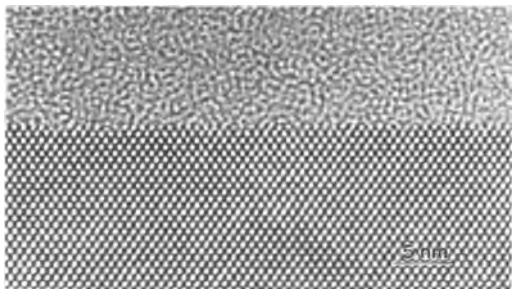


Figure 10. High-resolution image in ⟨110⟩ projection of the interface between thermal SiO_2 and (100) oriented silicon substrate.

detail by various analytical techniques. HREM of ⟨110⟩-oriented cross-sections was the first technique to reveal the (100)Si–SiO_2 interface on an atomic scale [27]. Figure 10 shows an HREM image in the ⟨110⟩ projection of a (100)Si–SiO_2 interface fabricated by a standard dry oxidation process. In the silicon crystal one set of {200} and two sets of {111} lattice fringes are visible. The dark dots in the crystalline part correspond to the position of two projected columns of silicon atoms being 0.136 nm apart, whereas the white dots are the tunnels in the crystal structure. The amorphous oxide shows a grainy contrast which does not reproduce the amorphous structure. Viewing the image parallel to the interface at an oblique angle reveals steps being one or two {200} lattice fringe spacings high. Characterization of irregular roughness by HREM of thin cross-sections suffers from the problem that even at very low specimen thickness (<10 nm) atomic elevations and depressions are projected on top of each other. Hence, simple visual inspection of the interface corrugation is bound to give the impression of too smooth interfaces. A quantitative approach by digitizing the image of the interface and calculating its covariance [28] yields values for the root mean square (r.m.s.) roughness and a correlation length. These parameters correspond to the peak-to-peak amplitude and the period of the roughness. The intentionally roughened (100)Si–SiO_2 interface oxidized in an ambient containing HCl was measured in this way [29], resulting in values of 0.3 and 2 nm for r.m.s. roughness and correlation length, respectively. This corrugation was found to explain well the decrease in carrier mobility in the inversion layer of a MOS transistor due to carrier scattering at the rough interface.

High-resolution microscopy of thermally oxidized, almost perfectly flat (100)Si surfaces, which were grown by molecular beam epitaxy (MBE), revealed a transition layer of 1 nm thickness at the (100)Si–SiO_2 interface with distinct crystalline contrast [30]. The contrast was found to be different in [011] and [0–11] projections and was interpreted as a thin layer of trimidite (high-temperature phase of quartz) epitaxially grown on (100)Si during thermal oxidation. It was argued that this crystalline layer might not be continuous on a less flat surface. In the ⟨100⟩ HREM images of thicker specimen areas a pronounced dot contrast with ⟨110⟩ spacings was found at the interface [31]. With the help of image simulation it was demonstrated that this contrast comes from the projection of small silicon pyramides with {111} facets which form at the (100)Si–SiO_2 interface.

An atomic model of the interface structure cannot be derived from HREM images only. X-ray photoelectron spectroscopy (XPS) measurements suggest some transition region between silicon and

SiO_2 since silicon atoms with one, two, and three bonds to oxygen were identified in addition to the tetrahedrally coordinated silicon atoms in the substrate as well as in the SiO_4^{2-} tetrahedra of the amorphous SiO_2 structure [32]. Information on the oxidation states of silicon with a spatial resolution at the atomic level may be obtained by using a 0.25 nm diameter electron probe in a scanning transmission electron microscope with a field emission gun for analyzing the near-edge structure of the Si $L_{2,3}$ ionization edge in an electron energy loss spectrometer [33]. The step height and their spacings on (1 0 0)Si–SiO_2 interfaces may also be evaluated from (1 0 0) plan view TEM specimens without removing the oxide [34]. In the [1 0 0] diffraction pattern, forbidden {0 1 1} reflections appear which are due to incomplete silicon unit cells terminating at the interface. From their diffracted intensity the mean step height and step separation may be extracted. More recently, STM and AFM have been widely applied to study the silicon surface after various chemical surface cleanings [35]. Good correlation between initial roughness and gate oxide breakdown was reported.

Apart from the gate oxide, thin dielectrics are also required for the cell capacitors in DRAMs. Reducing the size of the memory cell while maintaining its capacity was first achieved by reducing the thickness of the oxide. Since thermal oxides of thicknesses below 10 nm are difficult to fabricate with high yield and high reliability, the area of the capacitor was increased using different concepts: by fabricating trench capacitors, all the sidewalls of the trenches etched into the silicon substrate contribute to the capacitor area while keeping the required cell area small. Alternatively, the capacitor can be 'stacked' on top of the transistor using two polysilicon electrodes (stacked cell).

Various aspects during fabrication of the dielectrics for trench capacitors will be discussed in the following. While SEM cross-sections are well suited for the inspection of the shape of the trench after reactive ion etching, TEM cross-sections are required to image the detailed configuration of the trench dielectrics with sufficient resolution (about 0.5 nm in bright field images).

Figure 11 is a horizontal section through a trench capacitor of a 4 Mbit DRAM with a thermal oxide as the dielectric [36]. Reactive ion etching of the trench was followed by a wet chemical etching process to round the bottom of the trench. By this process, trenches were produced with sidewalls having a preferential orientation along {1 0 0} planes. The oxide is not of uniform thickness along the perimeter of the trench, due to differences in

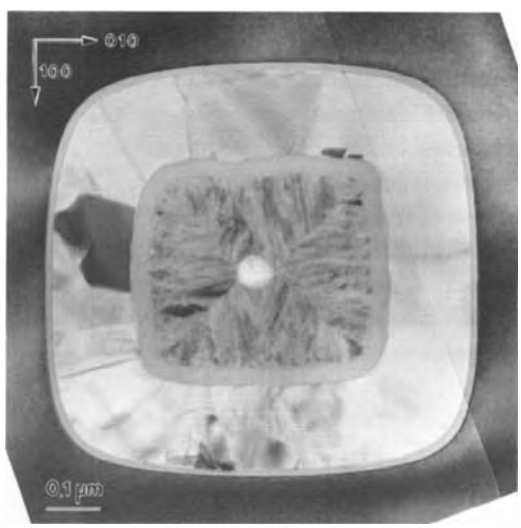


Figure 11. Horizontal section through a trench capacitor with a thin capacitor oxide.

the oxidation rate on the various crystallographic planes: oxidation on {1 1 0} planes is faster than on {1 0 0} planes. Whereas the oxide on the {1 0 0} sidewalls is approximately 13.5 nm thick, its thickness increases by about 30% to 17.5 nm in the rounded corners where the sidewalls are parallel to {1 1 0} planes.

The top and bottom edges of the trench are critical points with respect to thermal oxidation, because at temperatures below 1000°C the oxide thickness is reduced at the edges [37]. This effect always occurs at the upper edges of the trench, where patterning produces a sharp edge (Fig. 12a and b). Besides reducing the thickness of the oxide, this sharp-edged geometry impairs the breakdown strength of the oxide and forms a center of high mechanical stress which may nucleate dislocations [38]. At the bottom of the trench this problem can be avoided if well-rounded edges are produced by the etching process (Fig. 12c). In accordance with their different crystallographic orientations, the (1 0 0) trench bottom has a thinner oxide layer than the {1 1 0} trench sidewalls.

Because of the problem of oxide thinning at the upper trench edges, triple-layer dielectrics with an SiO_2–Si_3N_4–SiO_2 structure (ONO) are now employed as trench dielectrics. They benefit from the

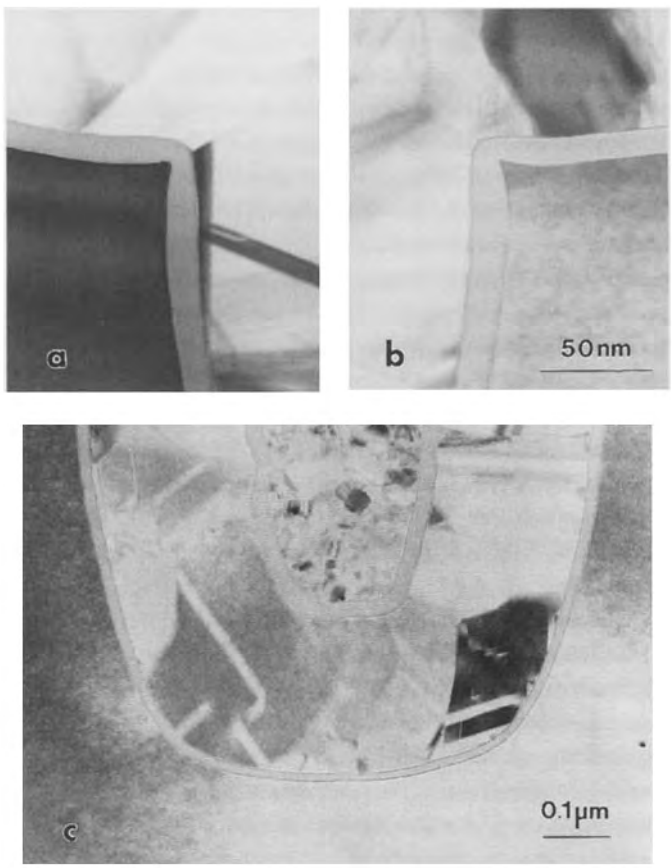


Figure 12. (a, b) Thinning of the thermally grown capacitor oxide at the top edges of a trench capacitor. (c) Cross-section of trench bottom. The center of the trench is filled with an additional thermal oxide and polysilicon.

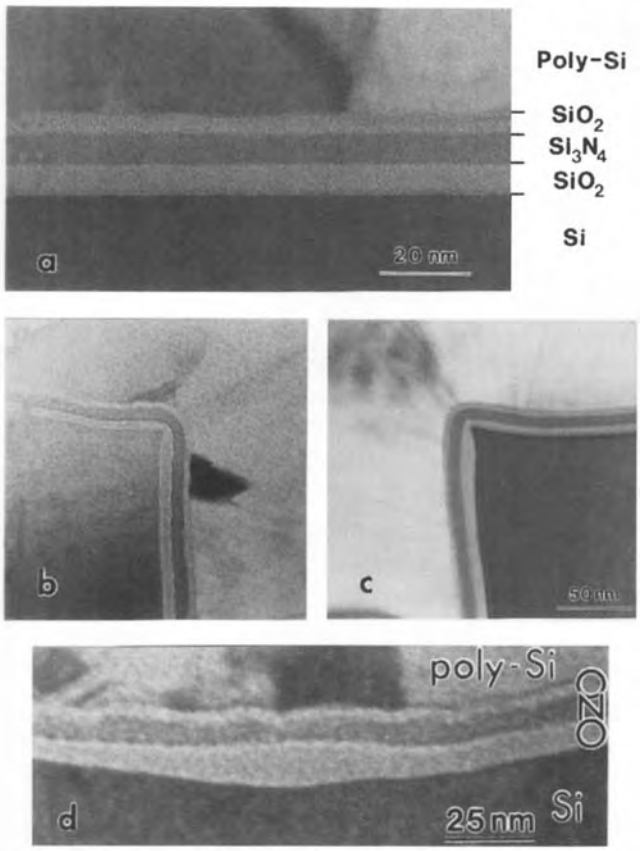


Figure 13. Bright field images of ONO triple layer dielectrics in the planar region (a) and at the upper trench edges (b, c). (d) Increased local oxidation of the bottom oxide at pores in the nitride layer.

fact that the nitride exhibits a higher dielectric constant ($\epsilon \approx 7$) as compared to SiO_2 ($\epsilon \approx 4$). Thus, reliable ONO dielectrics may be fabricated yielding the same capacitance as an effective ultrathin oxide. The oxide layers are needed to maintain well-defined Si–SiO_2 interface properties. The nitride of the ONO dielectric in Fig. 13a appears darker than the oxide layers because of its higher density resulting in enhanced scattering contrast. The configuration of an ONO dielectric at the upper edges of trenches is shown in Fig. 13b and c [36]. Again, the thermal bottom oxide is thinner at the edge. The nitride, in contrast, coats the edge with uniform thickness, because it was deposited from the vapor phase. The top oxide is best produced by thermal oxidation of the nitride. In this case, its thickness cannot be determined by ellipsometry, and TEM cross-sections represent the only means of measurement. The upper interface of the nitride layer is often slightly rough (Fig. 13d). This roughness is related to the nucleation of Si_3N_4 islands during the initial stage of deposition and may lead to the formation of pores in the nitride during oxidation. Since the oxidation rate of silicon is much higher than that of Si_3N_4, the bottom oxide tends to form lens-shaped areas at the pores (Fig. 13d) [39]. Besides the uniform step coverage of the nitride at the edges, this self-healing

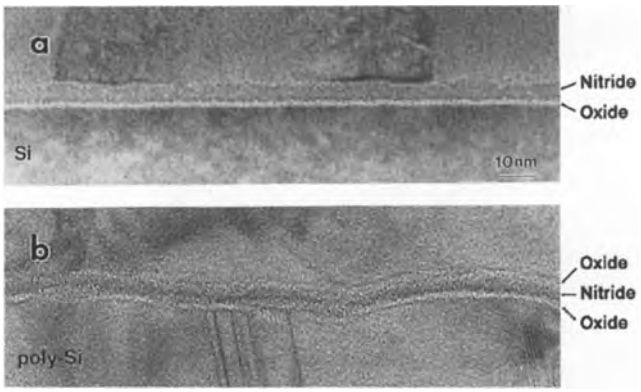


Figure 14. (a) Thin nitride/oxide layer on silicon substrate. (b) ONO layer on a rough polysilicon electrode.

effect, that is, the repair of local defects in the nitride during thermal oxidation, represents another advantage of ONO dielectrics.

For the capacitors in 16 Mbit and 64 Mbit DRAMs, progressively lower thicknesses of the dielectric are required. For the 64 Mbit DRAM, thicknesses will be as low as about 2 nm for the oxides and 5 nm for the nitride. Figure 14a shows an SiO_2–Si_3N_4 film on a silicon substrate. The nitride again shows a fine surface roughness. If such dielectrics lie between two polysilicon electrodes, the interfaces will not be planar, and TEM imaging in projection is only possible in limited regions (Fig. 14b) [39]. Recently, oxynitrides have drawn attention as possible gate dielectrics for submicrometer devices because they act as diffusion barriers for dopants from the polysilicon gate into the channel and have a low density of oxide charges and interface traps [40]. Future memory devices will employ novel perovskite-related materials, such as barium strontium titanate (BST) providing extremely high dielectric constants (up to 1000) or strontium bismuth tantalate (SBT) whose ferroelectric properties can be exploited for nonvolatile memory devices [41].

Using a high-intensity electron beam in the analysis of oxide and nitride layers may produce problems during TEM and also AES analysis. High-dose irradiation has shown that two effects occur with thermal oxides under the electron beam in the transmission electron microscope: (i) swelling of the oxide if there is no covering layer [42] and (ii) a thickness reduction if the oxide is covered by a polycrystalline silicon layer [43]. AES depth profiling of ONO layers with high electron current density leads to a pile-up of nitrogen at the Si–SiO_2 interface, whereas electron irradiation during TEM imaging of thin cross-sections induces nitrogen diffusion from the nitride layer to both Si–SiO_2 and polysilicon–SiO_2 interfaces [44]. Also ion beam milling may cause atomic mixing within the ONO layers [45].

3.3.2 Polysilicon and Metallizations

Modern CMOS circuits such as memory devices generally require three polysilicon levels for the gate and capacitor electrodes and for the first level of interconnections.

Since the conductivity of polysilicon is limited, primarily by the solubility limit of dopants in silicon, double layers of a refractory metal silicide and polysilicon—so-called polycides—are used for the gate and the first interconnection level. For additional interconnections, three or more metallization levels are required. Layers of aluminum alloys are used for this purpose, sandwiched between barrier layers, preventing interdiffusion at the contact interfaces and serving as anti-reflective coating for lithography. Since the conductivity of the interconnection layers depends strongly on their microstructure, and contact resistance is influenced by interface reactions, TEM cross-sections are used predominantly in the following to describe various materials effects.

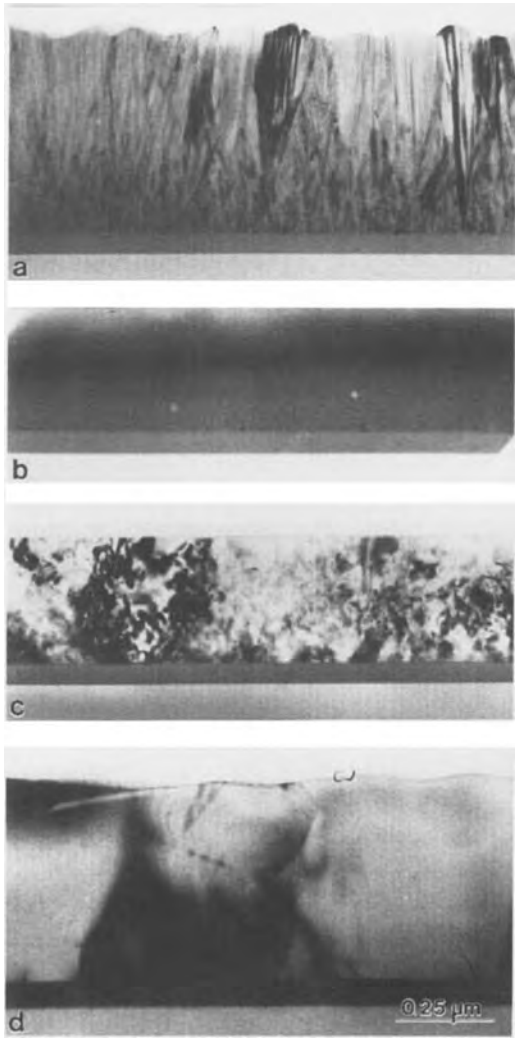


Figure 15. TEM cross-sections of polysilicon films deposited by LPCVD on SiO_2. (a) Crystalline deposition at 620°C. (b) Amorphous deposition at 560°C. (c) Crystallization during annealing (950°C) of an amorphously deposited film. (d) Grain growth due to high arsenic doping.

3.3.2.1 Polysilicon

Polycrystalline silicon (polysilicon) films are fabricated by LPCVD (low-pressure chemical vapor deposition) through the thermal decomposition of silane at temperatures between 550 and 630°C. Polysilicon technology is described in detail by Kamins [46]. The microstructure of polysilicon is strongly influenced by deposition and doping parameters; an overview is given by Cerva and Oppolzer [47]. Deposition at 620°C results in the growth of polycrystalline films (Fig. 15a) with a grain size of a few nanometers at the bottom interface. Conical grains start to grow at this interface with a cone angle of about 20°. Consequently, their grain diameter parallel to the film plane increases almost linearly from the bottom to the top interface and reaches about 120 nm at the top of a 0.5 µm thick film. The conical grains contain a high density of microtwins having a width of only a few nanometers. At the film surface, the microroughness of about 10 nm differences in height is determined by the grain structure,

with each grain forming a small hillock. Analysis by X-ray diffraction or RHEED (reflection high-energy electron diffraction) yields a preferred orientation of ⟨1 1 0⟩.

During deposition at 560°C, amorphous films are grown (Fig. 15b) with planar surfaces. At the bottom interface, small crystallites have nucleated. By an additional annealing step the amorphous film is crystallized which results in randomly oriented grains extending from the bottom to the top of the film (Fig. 15c). The film surface remains planar during crystallization. Since the silicon grains contain numerous microtwins the grain boundaries are not clearly visible in Fig. 15c, and grain size can best be determined by TEM dark-field images of plan view specimens. The shape of the grains is elliptical with the long axis parallel to a prominent microtwin running in the ⟨1 1 2⟩ direction. The growth of such grains may be explained by a specific mechanism [48]. Under standard annealing conditions of 800–900°C, crystallization takes place during heating up to the annealing temperature, since crystallization starts below 600°C. Therefore, the grain size depends on the process details, and larger grains grow at lower temperatures due to a reduced number of nuclei. Once crystallization is complete, no further grain growth occurs up to temperatures of about 1000°C. This almost holds also for polycrystalline deposited films: only the very small grains at the film bottom start to grow at 1000°C.

For application as gate or interconnect material the polysilicon films must be doped either by ion implantation, from the gas phase ($POCl_3$) or by the addition of doping gases during deposition (in situ doping). To obtain low sheet resistances, high doping levels up to the solubility limit (around $10^{20}\,cm^{-3}$) have to be applied. Boron doping has only a minor effect on the microstructure of polysilicon films. Large grains are only obtained by doping the amorphously deposited film prior to crystallization. Arsenic and phosphorus doping to produce highly n-doped polysilicon, on the other hand, causes strong grain growth, especially in polycrystalline deposited films. Grain growth stops when the grain size is of the order of the film thickness (Fig. 15d). All defects within the grains, that is, mostly the microtwins, disappear. When the doping level and/or the annealing temperature is not sufficiently high to cause pronounced grain growth the sheet resistance is especially sensitive to the microstructure, since the dopant atoms may segregate to the grain boundaries where their electrical activity is reduced. Strong segregation in combination with small grain size, that is, a large area of grain boundaries, results in unexpected high sheet resistances [49].

The segregation of arsenic to grain boundaries in highly arsenic-doped samples was measured directly by X-ray microanalysis in a scanning transmission electron microscope (STEM) equipped with a field emission gun [50, 51]. In a sample implanted with a dose of $2 \times 10^{16}\,cm^{-2}$ and annealed at 950°C for 30 min, arsenic profiles were recorded across grain boundaries which were parallel to the electron beam. Such a grain boundary is shown in Fig. 16a between A and B. The profile in Fig. 16b was obtained with a distance between the measurement points of 5 nm, and shows that the arsenic concentration is enhanced at the grain boundary by a factor of

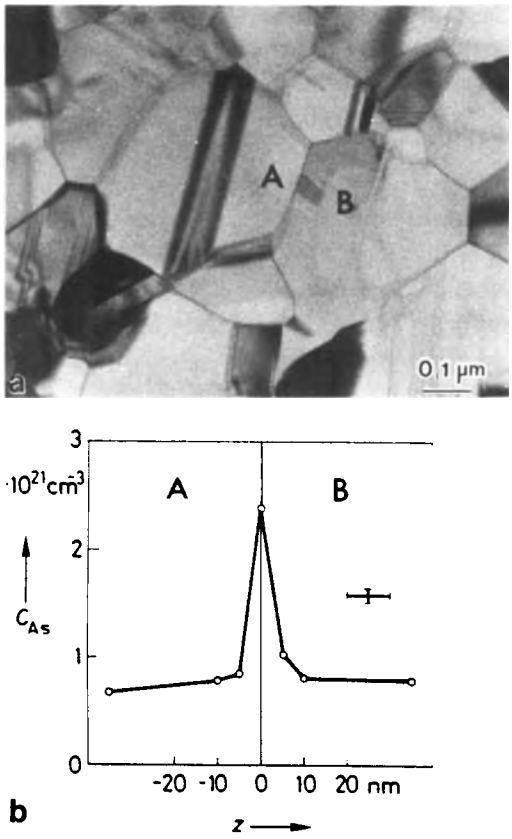


Figure 16. Arsenic segregation to grain boundaries in polysilicon. (a) Bright field image of the microstructure. (b) Arsenic concentration profile across the grain boundary between A and B in (a) measured by X-ray microanalysis with a 1 nm electron probe in a field emission STEM.

about 3. The arsenic concentration at 5 nm distance from the grain boundary is slightly higher than in the grain interior. Since this is due to beam spreading of the probing electron beam the width of the zone with increased arsenic concentration must be smaller than 5 nm, probably only a few monolayers [51]. The amount of arsenic atoms segregated to the grain boundaries was found to correspond approximately to one monolayer. From the grain size of 0.3 μm and the mean arsenic concentration of $8 \times 10^{20}\,cm^{-3}$, it follows that about 30% of the total arsenic atoms are segregated to the grain boundaries.

3.3.2.2 Refractory Metal Silicides

Double layers of polysilicon and silicides provide sheet resistances of only about one-tenth that of highly doped polysilicon single layers. The silicides of refractory metals such as titanium, molybdenum, tungsten, and tantalum have the advantage of being compatible with standard process technology: they are stable at temperatures as high as 900°C and are resistant to chemical wafer cleaning and etching solutions containing HF. The polysilicon is still needed along with the silicide, firstly because the well-defined interface with the SiO_2 is retained and secondly because it acts as a silicon source or drain during the annealing process, thus assuring reproducible manufacturing of stoichiometric disilicides. An overview on the application of silicides in silicon technology was given by Murarka [52].

The polycide of Fig. 17 contains a tantalum disilicide layer which appears dark due to increased scattering contrast caused by the high atomic number of tantalum. The polycide is shown after sputter deposition, annealing to form crystalline $TaSi_2$, and thermal oxidation. The roughness of the $TaSi_2$–polysilicon interface is determined by the globular shape of the $TaSi_2$ grains and indicates some interface reaction. During thermal oxidation an SiO_2 film grows on top of the polycide. Silicon diffusion through the $TaSi_2$ layer is so fast that it does not restrict the oxidation rate. The oxide film contains

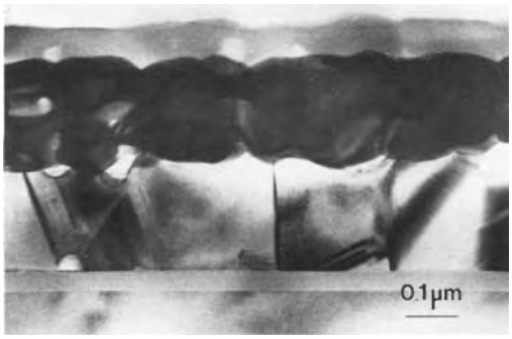


Figure 17. TEM cross-section of $TaSi_2$-polysilicon layer after thermal oxidation producing an SiO_2 layer on top.

numerous pores which are assumed to be responsible for the poor breakdown strength of such oxides as compared to SiO_2 on (1 0 0) silicon, like gate oxides [53].

In the case of a specific sputtering process for molybdenum silicide, a fine columnar structure was found in the amorphous material only at the steep sidewalls of contact holes (Fig. 18a) [39]. The bright lines represent low-density material forming channels with a honeycomb structure. If annealing (900°C) is performed in an oxidizing ambient, oxidation proceeds fast along the fine channels between the columns (Fig. 18b). The $MoSi_2$ no longer forms a continuous film, since the individual grains are surrounded by SiO_2. It is obvious that this effect, which does not occur in the planar film regions, increases the sheet resistance of such interconnections crossing contact holes or surface steps drastically. By changing the annealing conditions such that the initial annealing is performed in nitrogen, $MoSi_2$ films with a continuous microstructure and low sheet resistance were obtained.

In the 'salicide' process, a self-aligned silicide is grown in a contact hole, for example, without additional patterning

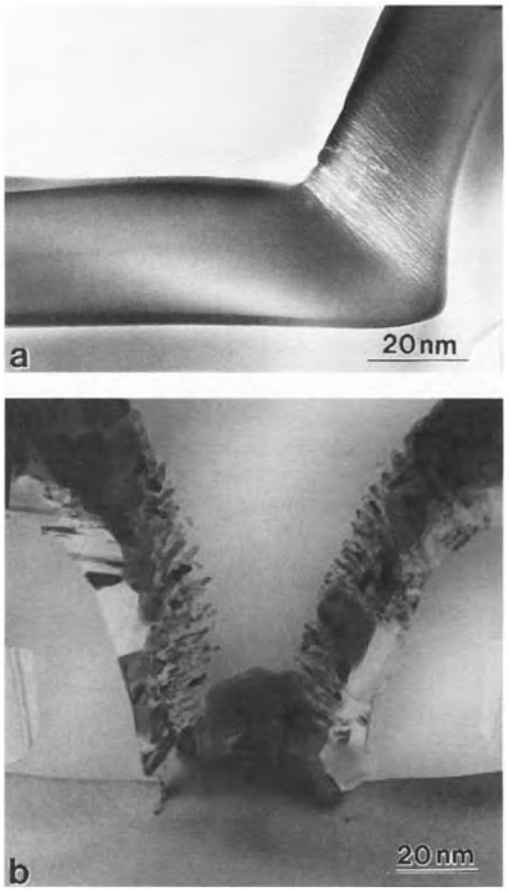


Figure 18. (a) Amorphous molybdenum silicide in a contact hole after sputtering. (b) Annealing (900°C) in an oxidizing ambient leads to a discontinuous $MoSi_2$ film at the sidewalls.

[52]. This process starts with depositing a thin metal layer on top of the patterned oxide on the silicon wafer. During annealing at a moderate temperature (<700°C), the metal reacts with the silicon substrate only in the oxide windows. When the metal which has not reacted is removed from the oxide by selective wet chemical etching, a silicide is left that has grown only in the oxide windows. Since the silicide–silicon interface was produced by a chemical reaction, it is free of contaminants and

has a low contact resistance. If the salicide process is applied to an MOS transistor structure in which the sidewalls of the polysilicon gate are masked by oxide spacers, a silicide is grown on the polysilicon at the same time as growth proceeds in the source/drain regions. As a result, sheet resistivity is reduced both in the polycide and in the source/drain regions. Figure 19a shows such a transistor structure with a titanium 'salicide'. Titanium silicide exhibits the highest thermal stability [52] but requires two-stage annealing. During the first step under a nitrogen atmosphere at a temperature below 700°C, the selective reaction produces the high-resistive C49 phase of $TiSi_2$. Following selective etching of the reacted TiN, the second annealing step is carried out at a higher temperature (e.g., 900°C) to form the low-resistive $TiSi_2$ of the desired phase (C54) [54].

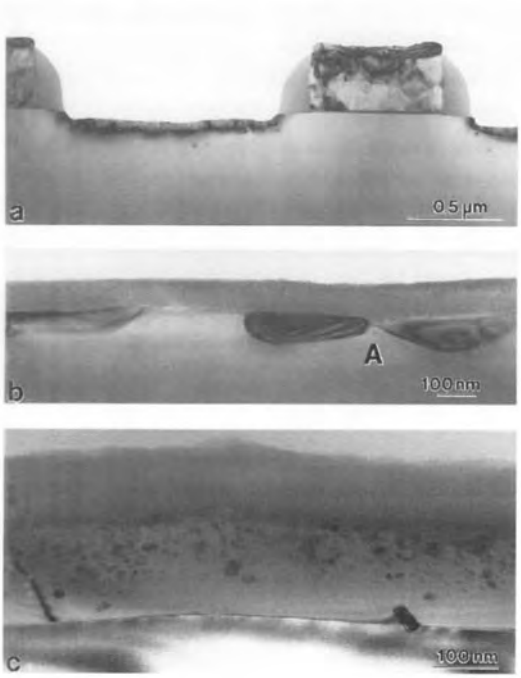


Figure 19. TEM cross-section of $TiSi_2$ films ('salicide' process). (a) MOS transitor with $TiSi_2$ on both the polysilicon gate and the source/drain regions. (b) Breaking up of $TiSi_2$ layer after high-temperature annealing. (c) TiB_2 precipitates at the interface to the substrate after boron implantation ($10^{16}\,cm^{-2}$) and annealing (950°C).

Prolonged annealing at 900°C, necessary for the reflow of borophosphosilicate glass (BPSG) to achieve planarization, however, increases the sheet resistance of $TiSi_2$ layers. This increase is due to morphological degradation of the $TiSi_2$ layers. Irregularities in film thickness due to nonuniform reaction are enhanced. Especially at grain boundaries the film thickness decreases strongly (at A in Fig. 19b) since the grains tend to form a globular shape in order to reduce their surface energy. In the end this process leads to breaking up of the silicide layer with the holes filled by epitaxially regrown silicon (Fig. 19b). A titanium salicide process with increased titanium thickness and an in situ cleaning step by back-sputtering prior to metal deposition proved to withstand the temperature of a BPSG reflow step (900°C for 40 min) without severe morphological degradation [55].

Apart from the decrease in sheet resistance, the contact resistance of $TiSi_2$ to highly doped silicon also increases during high-temperature treatment. This is due to a decrease in dopant concentration at the interface since the dopants diffuse from the silicon into the undoped silicide, where they are easily distributed because of their high diffusion coefficient in the silicide. Dopants in the silicides are also trapped by the formation of metal–dopant compounds. A detailed study [56] showed that in the case of boron doping, TiB_2 precipitates are formed at the interface (Fig. 19c). Their phase could be identified

by both lattice imaging (HREM) and X-ray diffraction after selective removal of the silicide leaving the precipitates at the surface. In the case of contacts to arsenic-doped silicon, TiAs precipitates are formed. To keep the increase in contact resistance at an acceptable low level, the thermal budget after silicide formation has to be minimized.

3.3.2.3 Aluminum Metallization

In the past aluminum–silicon alloys were the most widely used metallization system in silicon technology [52]. The addition of a weight fraction of about 1% silicon inhibits the dissolution of silicon from the substrate within the contact hole into the aluminum layer during annealing, a process that is usually carried out at temperatures around 450°C. The solid solubility of silicon in aluminum at 450°C is about 0.8% by weight fraction. If the silicon content is below the solubility limit because of local silicon depletion or excessive temperatures, interdiffusion takes place during contact sintering: silicon diffuses from the substrate into the aluminum and vice versa. A feature known as an aluminum spike then is formed in the contact hole. If all the technology layers are removed by etching, the spikes can be seen as pyramid-like pits in the silicon substrate (Fig. 20a). The formation of spikes results in pn junction leakage or shorting. On the other hand, the excess silicon in the Al (1% Si) metallization precipitates during cooling and forms silicon particles. Silicon precipitation may also take place in the contact hole, in which case the silicon grows epitaxially on the substrate. In Fig. 20b most of the

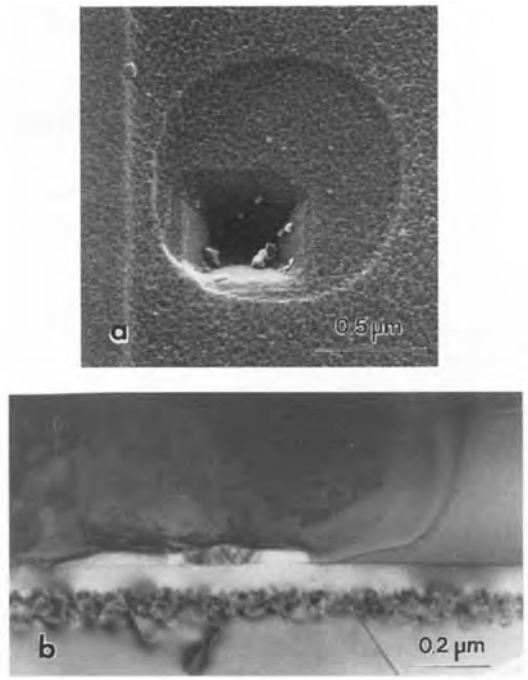


Figure 20. (a) Aluminum spike in a contact hole imaged in the SEM after etching off all device layers. (b) TEM cross-section of a contact hole with epitaxial silicon islands in an Al(1% Si) contact to the silicon substrate.

contact hole is covered by epitaxial silicon. The interfacial oxide layer is broken up into small oxide particles, which make the original interface show up weakly as a dark line. The epitaxial islands contain microtwins. Since the epitaxial silicon precipitates from the aluminum, it is p conductive and produces regions with nonohmic contact behavior for contacts to n-type silicon. The contact resistance of contacts to p-type silicon is increased. Both problems, spiking and epitaxial precipitation, can be overcome by barrier layers.

With shrinking device dimensions the width of both interconnect lines and contact holes decreases. The increased current density in narrower aluminum lines

enhances the risk of failure due to electromigration, that is, current-induced material transport leading eventually to the interruption of lines. Adding about 0.5% copper, for example, to aluminum minimizes this risk but requires the use of additional diffusion barriers, since copper contamination in the active device regions reduces the lifetime of charge carriers and generates crystal defects causing leakage currents. As mentioned above, diffusion barriers are also required to inhibit interdiffusion effects such as spiking and epitaxial regrowth, which are critical for small contact holes in particular. On account of its high thermal stability and good electrical conductivity, TiN is one of the best suited materials. To improve adhesion on SiO_2 and obtain a low contact resistance on p- and n-type silicon, a thin layer of pure titanium is deposited before the TiN. Both layers are deposited in one run with the TiN layer being produced by reactive sputtering from a titanium target in an $Ar–N_2$ gas atmosphere. The composition of the TiN_x layer is controlled by the $Ar–N_2$ flow ratio. Since only TiN_x layers with $x \approx 1$ exhibit favorable properties, their composition must be analyzed quantitatively with high accuracy. Based on calibration by RBS (Rutherford backscattering spectroscopy), this can be achieved by AES. AES with a small electron probe (<0.1 µm diameter) even allows local analysis of the TiN_x composition, for example at the sidewalls of contact holes.

For investigation of the thermal stability of the Ti–TiN barrier, specimens annealed at 450 and 550°C were examined by TEM (Fig. 21). The TiN has a pronounced columnar microstructure. Crystallites with a diameter of 5–10 nm extend almost throughout the entire film. The bright lines indicate voids or narrow regions of low density at the grain boundaries which, however, are probably 'stuffed' with oxygen due to exposure to air after film deposition. Auger analysis showed the TiN to indeed contain several

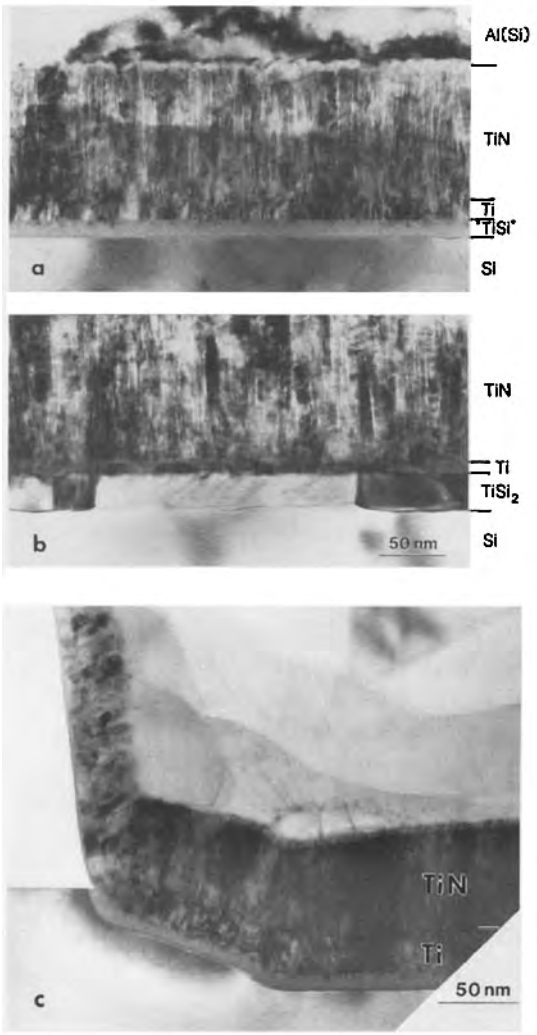


Figure 21. (a, b) Ti/TiN barrier layer between aluminum metallization and silicon substrate after annealing at 450°C (a) and 550°C (b). Ti/TiN layer at the bottom corner of a contact hole with reduced thicknesses, especially at the sidewall.

percent of oxygen. The titanium layer reacted with the silicon substrate in both specimens. Whereas the reaction zone is amorphous in the specimen annealed at 450°C (Fig. 21a), it is crystalline with grains some 100 nm in size after annealing at 550°C (Fig. 21b). The composition of this zone was measured by X-ray microanalysis of the thin cross-sections using a scanning transmission electron microscope equipped with a field emission gun. The measurements revealed that the amorphous layer has a composition close to TiSi, whereas the crystalline layer consists of $TiSi_2$ [36].

While the contact holes become smaller, the thickness of the insulating dielectric layers remains largely constant. Hence, small contact holes have a high aspect ratio, defined as the ratio of depth to diameter. To ensure reliable contacts, a certain step coverage much be achieved during deposition of the metallization. Step coverage is the ratio of layer thickness on a vertical sidewall to the nominal layer thickness, and is a function of the aspect ratio and the deposition process. Sputter deposition suffers from shadowing effects that reduce film thickness at the bottom and on the sidewalls of the contact hole. Figure 21c shows a Ti–TiN barrier at the bottom corner of a small contact hole. The titanium and TiN thicknesses of 50 and 100 nm besides the contact are reduced to 8 and 23 nm at the sidewall, and 11 and 28 nm at the bottom of the hole, respectively, not taking into account the amorphous reaction zones. At the sidewall the TiN layer can be distinguished from the titanium by its columnar microstructure. Besides this poor step coverage, the detailed shape of the sidewall is critical since protruding parts, such as of an additional nitride layer, lead to interruptions of the barrier and possibly allow direct contact between the aluminum alloy and silicon. Collimated sputtering of TiN increases the barrier thickness at the bottom of the contact hole, and CVD deposition provides uniform coverage.

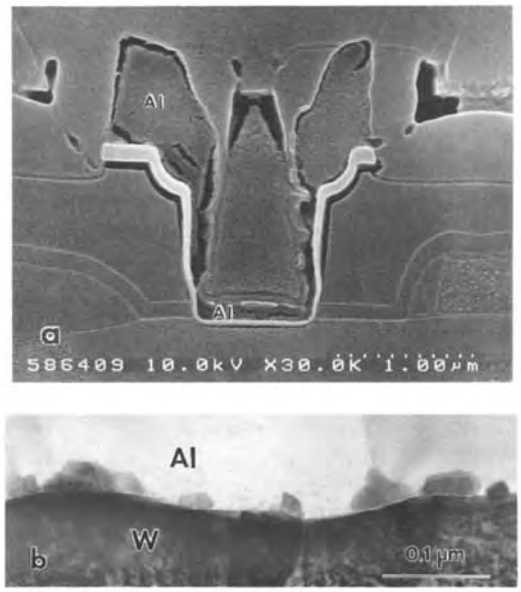


Figure 22. (a) SEM cross-section of a contact hole showing poor step coverage of the aluminum film. (b) Grains of the additional phase $W(Si,Al)_2$ at the tungsten–aluminum interface.

Also, the sputter deposition of the aluminum metallizations suffers from poor step coverage. The SEM cross-section of Fig. 22a shows that the aluminum thickness of 0.8 µm outside of the contact hole is reduced to only 0.1 µm in the contact hole bottom. At the sidewalls the thickness is locally even lower. When the step coverage becomes unacceptably low, the problem can be solved by a different approach, that is, filling the contact holes with tungsten plugs. This requires a two-stage process. First, tungsten is deposited uniformly by CVD on the patterned wafer covered by

a Ti–TiN barrier layer. Second, the tungsten in the planar regions is etched back until the etch-back stops at the TiN layer. When the tungsten film thickness is sufficiently high, the contact holes remain well filled by tungsten after the etch-back. Such tungsten plugs are used for both contacts to silicon and for 'vias', that are contacts between different aluminum metallization levels. In Fig. 1 such a tungsten plug can be seen between two gates, appearing as a bright needle because of its high atomic number. At the interface between tungsten and an AlSiCu metallization (1% Si and 0.5% Cu) a reaction was observed after annealing at 450°C [57]. Grains of an additional phase are formed at the interface within the AlSiCu layer (Fig. 22b). These grains—about 30 nm in size—were identified by microdiffraction in the transmission electron microscope to consist of hexagonal $W(Si,Al)_2$. No detrimental effect on contact resistance, however, was observed.

3.3.3. Lithography and Etching

In combination with the fabrication of dielectric and conductive layers and forming the active device areas by doping, generation of the device patterns is the main task in IC manufacturing. Patterning comprises both the generation of a resist pattern on the wafer by optical lithography and transferring this pattern into the underlying layer by an etching process. During fabrication of ICs, patterning is performed many times for the various regions and layers such as implantation regions, contact windows, interconnections, and so forth. In lithography the photoresist is exposed with the pattern of the mask using a projection system reducing the mask (reticle) size by a factor of usually 5. The whole wafer is exposed in a 'stepping' procedure. After development the resist pattern resembles that of the mask (reticle) for a positive resist. Using deep ultraviolet light, optical lithography is able to generate patterns with dimensions below 0.2 μm and therefore will dominate in the near future. Dry etching is predominantly employed for pattern transfer since its high anisotropy produces steep sidewalls. The proper choice of reactive ions in combination with optimized etching parameters provides high selectivity when the plasma-assisted etching process reaches the underlying material.

To solve specific problems during the development of lithography or etching processes, the detailed information provided by cross-sections is frequently necessary to measure features, such as step coverages or the exact shape and angle of sidewalls. Since the preparation of cross-sections by cleavage fracture or polishing destroys the processed wafer, this type of SEM characterization is performed outside the fabrication line. The detailed procedures, such as slight selective etching of the cross-sections to obtain a better contrast between the various layers, are identical to those used in failure analysis, and will be discussed in Sec. 3.4.2 of this Chapter. Process control of lithographic and etching processes, however, requires nondestructive assessment within the fabrication line with a short turnaround time. This topic will be discussed separately in Sec. 3.4.1 of this Chapter.

After reactive ion etching (RIE), residual surface films can be formed on both the sidewalls and at the bottom of contact

holes. During RIE of a $TaSi_2$–polysilicon (polycide) layer using chlorine-based chemicals a film of about 20 nm thickness was found at the sidewalls [58]. Anisotropic etching during RIE is assumed to be due to protective polymer films at the sidewalls stemming from residues of the photoresist. Cross-sectional TEM and AES, however, showed this film to consist of SiO_2 sputtered from the oxide lying under the etched polycide. Sputtering of the oxide takes place after the polycide is etched through.

After etching of contact holes in CVD oxides, thin insulating layers may be present at the bottom of the contact holes, degrading the contact resistance seriously. Figure 23a shows a TEM cross-section of an aluminum contact to polysilicon. A thin layer of about 5 nm thickness is clearly seen between the silicon substrate and the polysilicon (Fig. 23b). From the very high resistance of this contact the layer was assumed to consist of SiO_2. Because RIE is assisted by ion bombardment, a considerable amount of lattice damage is introduced in the underlying material once the contact hole is etched through, and native oxide growth is enhanced when the etched wafer is brought back into the air. To avoid excessive oxide growth, the etching process must be optimized by appropriate post-treatment steps in order to remove the damage zone without introducing new damage. During RIE process development the thickness of the thin oxide layer at the bottom of contact holes had to be determined with a short turnaround time. Since cross-sectional TEM is too elaborate for this purpose, a method was developed [59] based on AES which provides both sufficient lateral resolution and surface

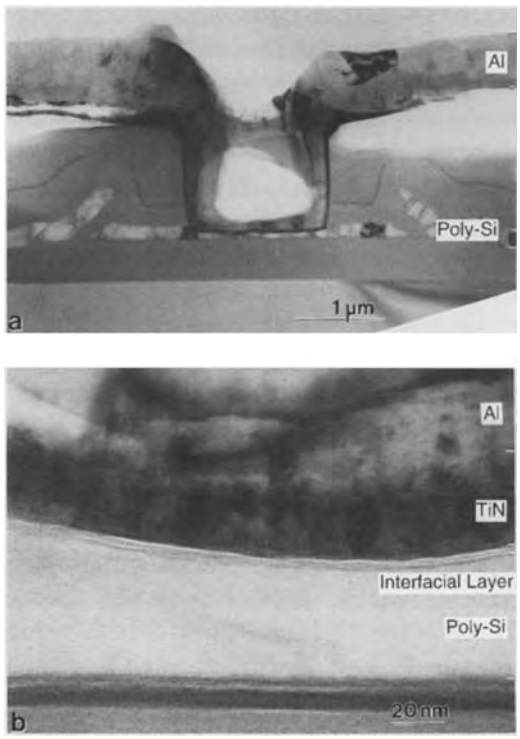


Figure 23. (a) TEM cross-section through an aluminum contact with high contact resistance to a polysilicon interconnection. (b) Higher magnification of the contact region revealing a 5 nm thick interfacial layer.

sensitivity. Application of AES to this problem of contact holes, however, is not straightforward. Conventional sputter depth profiling cannot be applied in such narrow holes because material sputtered off from the sidewalls will be redeposited at the bottom. To avoid any sputtering, the Auger electron lineshape was utilized to monitor the film thickness, which was below 10 nm. The peak shapes of the silicon KLL Auger transitions differ in a characteristic way between elemental and oxidized silicon (Fig. 24a and b). Since the Auger spectrum of the contact hole is a superposition of the spectra from bulk silicon and SiO_2, it can be fitted with the

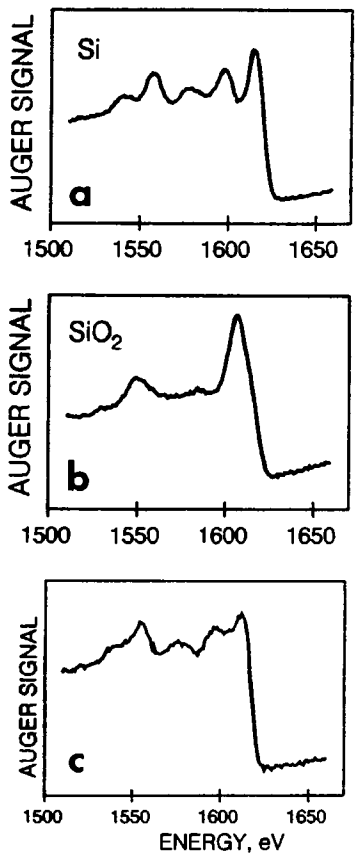


Figure 24. (a, b) Auger spectra (silicon KLL transition) of silicon and SiO_2. (c) Auger spectrum of a thin oxide film in a contact hole representing a superposition of spectra (a) and (b).

respective reference spectra. The oxide thickness follows from the fitting parameters [59]. Figure 24c shows the spectrum from a thin oxide film in a contact hole, the thickness of which was determined to be 1.7 nm. Process optimization resulted in a reduction of the oxide thickness to a level of the native oxide which is not critical for contact failure. Shadowing of the Auger electrons emitted from the bottom of the contact hole by the sidewalls, however, restricts this method to contact holes with an aspect ratio of about 1. For higher aspect ratios, appropriate sample preparation is necessary, for example, by cleavage fracture through the holes.

3.3.4 Doping

To produce the active device areas, dopant atoms have to be introduced into the silicon substrate within selected regions defined by a mask. Compared to diffusion from the gas phase or doped-oxide sources, the most commonly used method of doping is ion implantation with boron for p doping and arsenic or phosphorus for n doping. Ion implantation allows the doping depth and concentration to be varied within wide limits by selecting the acceleration voltage and the dose. The wells for CMOS devices, for example, are only lightly doped ($\sim 10^{16}\,cm^{-3}$) and may extend up to several micrometers into the substrate. Highly doped regions ($>10^{19}\,cm^{-3}$) with pn junctions below 100 nm depth are required for the source and drain regions of MOS transistors or the emitter of bipolar transistors, for example. Since the nuclear collisions that occur during implantation damage the crystal lattice, the defects produced by this damage have to be annealed by heat treatment. Only after annealing are the dopant atoms located on substitutional sites and electrically active. Characterization of the lattice damage after ion implantation and of the residual defects after annealing will be discussed in Sec. 3.3.5 of this Chapter on process-induced defects. Since the electrical device parameters depend critically on the depth distribution of the dopant atoms, the dopant depth profiles have to be accurately

measured by SIMS. SIMS is the method of choice since it combines sufficient sensitivity (down to dopant concentrations of $10^{14}\,cm^{-3}$) with good depth resolution (down to <10 nm). An introduction to SIMS can be found in Fuchs et al. [5], and a more detailed description in Benninghoven et al. [60].

SIMS profiles are usually measured on test patterns, for example $100 \times 100\,\mu m$ in size. Although the primary ion beam is scanned across the silicon surface to provide uniform erosion, no lateral information is required in general. In specific cases, however, SIMS profiles have to be measured within actual device structures. This calls for spatially resolved SIMS (micro-area SIMS). As an example, the analysis of the dopant depth profile at the sidewalls of trench capacitors in 4 Mbit DRAMs [61] will be discussed. One electrode of this capacitor is formed by highly n-doped polysilicon and the other by the wall of the trench doped by diffusion from a CVD oxide which contains arsenic. Since systematic differences were observed between the capacity of the trench cells and analogous planar structures, indicating geometrical influences of the CVD process, it was important to determine the actual dopant depth distribution at the sidewalls of the trenches directly. For this purpose cross-sections were prepared (Fig. 25a) and analyzed by micro-area SIMS with a primary ion beam diameter of only 0.3 μm. To obtain high sensitivity the 'useful ion yield', that is, the ratio of detected ions and the atoms sputtered from the sample, must be optimized. In this case of arsenic in silicon an optimum useful yield of 2×10^{-2} was found for $AsSi^-$ secondary ions under cesium bombardment. During each scan of the

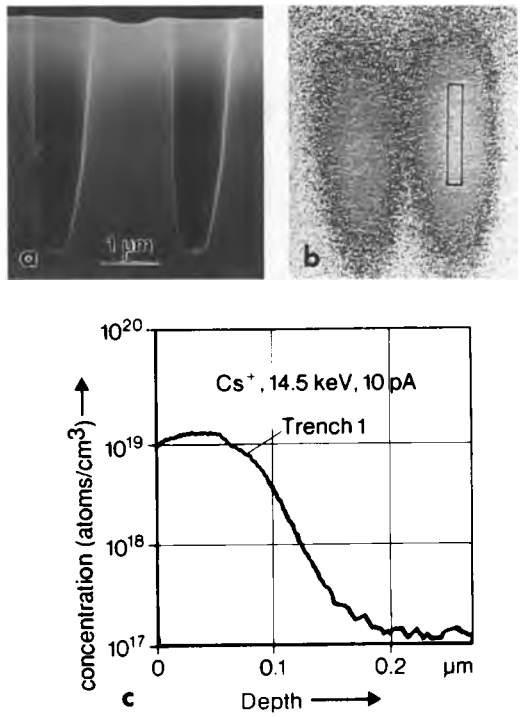


Figure 25. Micro-area SIMS of the arsenic depth profile at the sidewalls of trench capacitors in a 4 Mbit DRAM. (a) SEM image of a cross-section through two trenches. (b) Arsenic distribution using the $AsSi^-$ signal. (c) Arsenic depth profile from the selected area indicated in (b).

analyzed area the intensity of the $AsSi^-$ ions was stored in a digital image storage system, allowing the area of interest to be selected retrospectively from the stored data. The area that contributes to the profile can thus be optimally matched to the size of the trench wall. Figure 25b shows such an $AsSi^-$ image with the area selected for the profile. The resulting dopant profile is shown in Fig. 25c. Despite the high efficiency for secondary ion detection of the SIMS system used, only a limited number of dopant atoms are available in the analyzed microvolume. This restricts the detection limit to about $10^{17}\,cm^{-3}$ in this case. Compared to planar

structures, both maximum concentration and the dose (total area below the curve) of arsenic were found to be lower at the trench sidewall. This difference was thought to be due to the oxide acting as a diffusion source being thicker on the planar surface than in the trench.

With decreasing device dimensions the lateral extension of dopants below mask edges is becoming more important, calling for an accurate determination of 2D dopant profiles with a resolution of 10–50 nm and a sensitivity range of 10^{16}–$10^{21}\,cm^{-3}$. This places serious demands on techniques to be applied to the characterization of 2D profiles. The easiest way to obtain 2D information of doped regions is selective (concentration-dependent) etching of vertical cross-sections and imaging of the etched surface topography by SEM, TEM, or STM, and AFM. SEM cross-sections are most widely used for technology characterization, and selective etching in addition provides junction delineation. Most etching solutions are mixtures of HNO_3, which oxidizes silicon, and HF, which removes the oxide. The addition of acetic acid controls the etch rate. The cross-section of an MOS transistor structure in Fig. 26a was etched in a solution of 0.5% HF in HNO_3. The highly arsenic-doped source and drain regions are preferentially etched and are bordered by a sharp contour because of the steep concentration gradient typical for arsenic depth profiles. This contour represents an isoconcentration line, and by comparison to SIMS profiles was found to correspond to a concentration of about $10^{19}\,cm^{-3}$ when applied to TEM cross-sections [62]. The contours reach about 100 nm underneath the oxide spacers on both sides of the gate. The so-called lightly doped drain

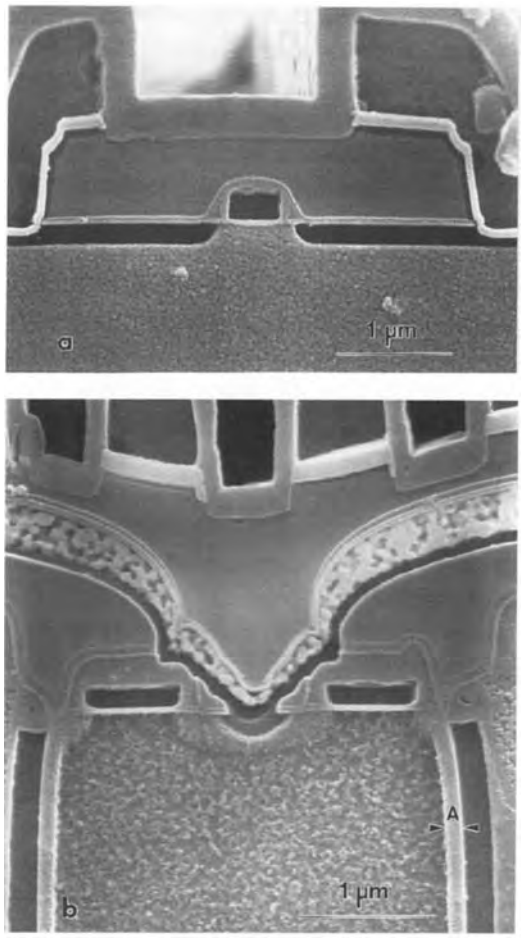


Figure 26. Delineation of pn junctions by selective etching of SEM cross-sections. (a) Highly arsenic doped source/drain regions of MOS transistor. (b) Bit line contact in a 16 Mbit DRAM.

(LDD) and source regions, which extend from the highly doped regions just below the gate, were not delineated. Since their maximum concentration is typically below $10^{18}\,cm^{-3}$ they were not attacked by this chemical etch. Figure 26b shows an SEM cross-section of the bit line contact of a 16 Mbit DRAM etched in a solution of 2% HF in HNO_3. The bit line consists of a polycide, the n-doped polysilicon of which is strongly etched and therefore appears

dark due to reduced secondary electron (SE) collection from regions lying deeper than the other surface. The highly arsenic doped contact region in the silicon substrate is also strongly etched. The source and drain regions extending from the bit line contact to the gates on both sides can be identified since they are attacked less than the silicon substrate despite higher doping. On the far side of both gates the doped sidewalls of the trench (region between the two arrows at A in Fig. 26b) are visible because of the same etching behavior. The highly n-doped polysilicon within the trenches is strongly attacked, similar to the gates. No quantitative comparison with SIMS profiles was made in this case. In Fig. 26a only one isoconcentration line was delineated. Other etching solutions yield up to three different isoconcentration lines for boron which can be quantified by SIMS [63].

When TEM cross-sections are selectively etched, these specimens are attacked from both sides. The solution of only 0.5% HF in HNO_3 [62] exhibits a reduced etching rate at 5°C [64]. When imaged in bright field under two-beam conditions, thickness fringes appear in the preferentially etched regions. They mark lines of equal specimen thickness and correspond to isoconcentration lines, provided the specimen has uniform thickness after ion milling prior to etching. In Fig. 27 five thickness fringes may clearly be discerned. The n-doped region was produced by arsenic implantation at an energy of 50 keV and a dose of $5 \times 10^{15}\,cm^{-2}$, followed by annealing at 900°C for 1 h [65]. Comparison with the SIMS depth profile showed the deepest thickness fringe to correspond to a concentration of about $5 \times 10^{17}\,cm^{-3}$. The contours corresponding to concentrations

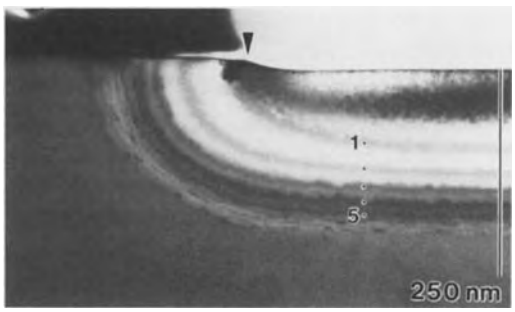


Figure 27. Thickness fringes in the n^+-doped region of a selectively etched TEM cross-section.

of 10^{19} and $10^{18}\,cm^{-3}$ at depths of 160 and 215 nm, respectively, were found to extend laterally 100 and 130 nm beyond the mask edge. The ratios of lateral extension and vertical depth are 0.63 and 0.61, respectively. Since the reproducibility of selective etching is limited, the accuracy and quantification of this method are restricted, and reference to SIMS profiles is necessary. Furthermore, when selective etching is applied to SEM cross-sections of actual device structures, identical transistors may yield different results if they are connected to different parts of the circuit having, for example, different capacities.

Various other microscopic techniques have been applied to characterize 2D dopant profiles [66], and were reviewed by Diebold et al. [67]. STM allows delineation of pn junctions by detecting differences in the tunneling characteristics for p-type and n-type semiconductor material [68]. Since the tunneling current depends on the Fermi level rather than on the carrier concentration, the application of this approach for profiling is restricted. Recently, a sharp conductive AFM tip was used to perform local resistance measurements between the tip and the back contact of the sample. When the tip is

scanned across a vertical cross-section the lateral carrier distribution can be determined [69], extending spreading resistance profiling (SRP) to nano-SRP. A metallized probe tip can also be used for scanning capacitance microscopy (SCM) when the tip is brought in contact with the oxidized semiconductor surface. The resulting MOS capacitance varies with carrier concentration. Recently, by calibration with SIMS, SCM was successfully applied to measure a 2D dopant profile quantitatively [70].

In the scanning electron microscope pn junctions can also be delineated by using the electron beam-induced current (EBIC) as a signal for imaging. When the electron beam strikes a pn junction the holes and electrons generated by the primary electrons are separated by the electric field of the pn junction, and a charge collection current, that is, the EBIC signal, can be recorded in a closed circuit. Only the minority charge carriers diffusing to the pn junction contribute to the EBIC signal when the electron beam generates electrons and holes at some distance from the junction. Therefore, the pn junction appears as a bright line, and its 2D shape is displayed directly when a cross-section is imaged. In the cross-section of a bipolar transistor in Fig. 28 all pn junctions between emitter (E), base (B), collector (C), and substrate (S) are visible [71]. In asymmetrically doped junctions (n^+p or p^+n) the EBIC maximum deviates considerably from the location of the junction. In addition, the finite size of the generation volume for charge carriers, which depends on the primary electron energy, limits the resolution of this method to above 0.1 μm. An introduction to junction delineation by EBIC is given by Fuchs et al. [5], and a more detailed review, also covering other applications of EBIC, by Leamy [72].

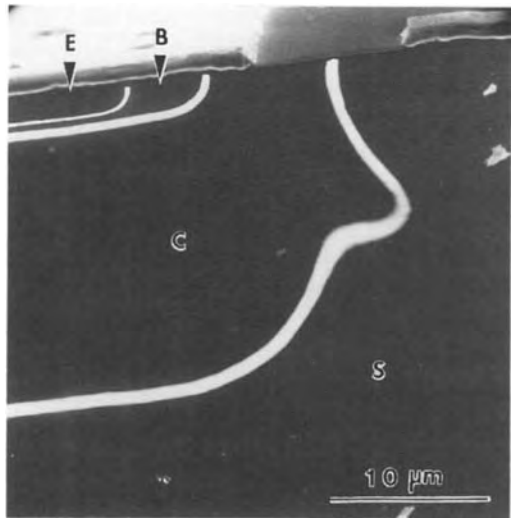
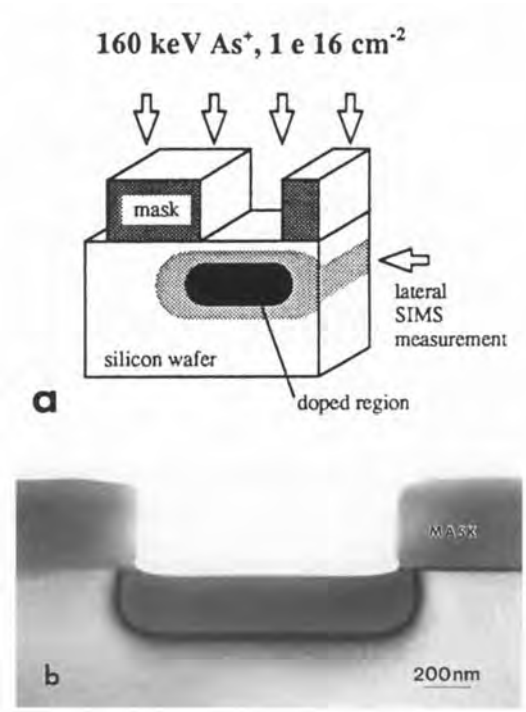


Figure 28. EBIC image (superimposed on SE image) of an integrated bipolar transistor showing all pn junctions between emitter (E), base (B), collector (C), and substrate (S).

The resolution and sensitivity of SIMS are not sufficient to measure the true 2D dopant distribution on a vertical cross-section directly. The lateral extension of doped regions underneath a mask edge, however, can be measured by SIMS, when a vertical cross-section parallel to the mask edge is analyzed as shown in Fig. 29a [73]. By selecting only the narrow, doped zone at the wafer surface for micro-area SIMS, the lateral dose distribution of the dopant atoms is measured with high lateral resolution. In this 'lateral SIMS' method the high lateral resolution is achieved by the good depth resolution (typically 10 nm) of SIMS profiling. Figure 29b shows the TEM cross-section of a test structure studied by lateral SIMS. A pattern of oxide stripes was implanted with 160 keV As^+ ions ($10^{16}\,cm^{-2}$ dose). The zone

amorphized by arsenic implantation appears dark (see Sec. 3.3.5). The lateral dopant distribution after ion implantation as measured by lateral SIMS is shown in Fig. 29c [74]. At a distance of about 50 nm from the mask edge, a maximum in the dose distribution is observed. To support the experimental results, Monte Carlo simulations of the implantation process were performed, taking into account the precise shape of the mask edge as displayed in Fig. 29b. The very good agreement is evident. The maximum in the dose distribution stems from ions which were implanted into the mask close to its edge but after several scattering events could leave the mask through the sidewall and reach the silicon substrate [74].



3.3.5 Process-Induced Defects

Crystallographic defects such as dislocations, stacking faults, and precipitates may form during particular steps in device processing, and have a direct influence on device performance and yield. Several reviews with a strong emphasis on TEM have been published [75–79]. There are several sources for defect generation in the various technology processes, and mutual interaction of the sources is possible.

Figure 29. 'Lateral-SIMS' method. (a) Principle: micro-area SIMS measurement on vertical cross-section. (b) TEM cross-section of test structure with oxide stripes and amorphized implanted regions. (c) Lateral dopant distribution with respect to mask edge.

3.3.5.1 Thermal Stresses

During high-temperature processes in a furnace the silicon wafers are positioned in a row with constant spacings. This leads to large radial temperature gradients between the wafer center and periphery because heat is dissipated differently during the cooling down procedure [75]. The temperature gradient increases with increasing wafer diameter, cooling rate, and furnace temperature. The stresses are highest at the wafer edge and cause dislocation generation when the stress exceeds the critical resolved shear stress. Generation of dislocations strongly depends on the availability of nucleation sources (e.g., microscratches, film edges, and precipitates). Impurities such as dissolved oxygen reduce the mobility of dislocations and impair their propagation. Precipitation of oxygen drastically reduces the yield strength and the critical resolved shear stress [75]. Hence, high-temperature steps after precipitation which cause large temperature gradients should not be applied. The stresses due to temperature gradients may be avoided by slowly inserting and withdrawing the wafers into/from the furnace at moderate temperatures (ramping procedures). Moreover, a minimum of grown-in bulk microdefects, or doping the silicon with nitrogen, reduces the generation of dislocations. X-ray topography allows nondestructive imaging of the distribution of stress-induced dislocations such as glide bands across the entire wafer.

3.3.5.2 Metal Precipitates

Heavy metals may contaminate the silicon wafer during processes such as wet cleans, dry etching, ion implantation, and furnace annealings, and through wafer handling and transport [80]. In particular iron, nickel, and copper are fast diffusers in silicon and have a relatively high solubility which declines steeply with decreasing temperature. Therefore, metal contamination on the wafer surface is dissolved at high temperatures and distributed in the silicon lattice. Upon cooling the metals precipitate as metal silicides when their concentration exceeds the solid solubility. Pre-existing defects—such as oxidation-induced stacking faults or implantation damage—and the $Si–SiO_2$ interface can act as heterogeneous nucleation sites for the precipitates, and dislocations may become decorated and electrically active. Precipitates at the gate oxide or the capacitor oxide reduce the dielectric breakdown field [81]. TEM investigations on intentionally contaminated silicon wafers revealed the different precipitation behaviors of copper and nickel and thus can explain their different impacts on oxide integrity [82, 83]. Nickel precipitates coherently as $NiSi_2$ with the CaF_2 structure, causing almost no volume expansion in the silicon lattice. Figure 30a shows a thin and long $NiSi_2$ platelet at the $Si–SiO_2$ interface penetrating into the oxide only slightly. The enlarged view in the high-resolution image (Fig. 30b) reveals that the $NiSi_2$ lattice is in a twin orientation to the silicon lattice and the $NiSi_2–Si–SiO_2$ interface exhibits notches. The cusp of the $NiSi_2$ platelet might alter the strength of the electric field across the oxide and lead to oxide failure. Copper silicides reduce the oxide breakdown field much more strongly than $NiSi_2$ [81]. According to structural and EDX analysis in TEM, copper silicide precipitates are

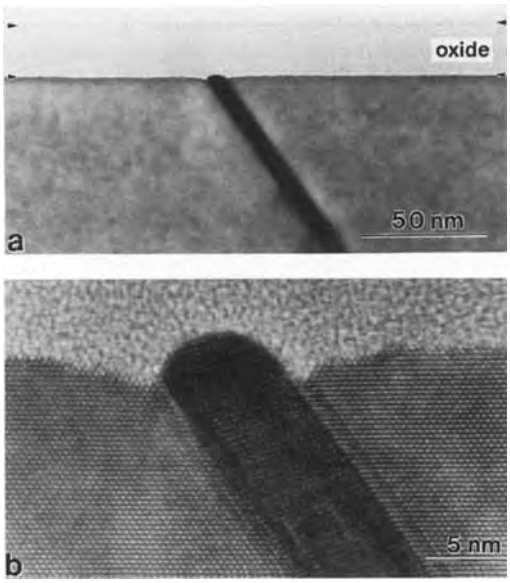


Figure 30. Precipitation of a $NiSi_2$-micrometer-platelet at the (1 0 0)Si-SiO_2 interface. (a) Bright field image of the top portion of a several-micrometer-long precipitate lying on a {1 1 1} plane. (b) HREM image of the top of the precipitate penetrating the oxide layer only slightly.

reported to be copper-rich [82, 84]. In any case, the volume required by the copper silicide precipitate contains a lower silicon concentration than the same silicon volume. Hence, copper silicide precipitation causes volume expansion and is accompanied by silicon interstitial emission or vacancy absorption. Figure 31a shows a copper silicide precipitate at the Si–SiO_2 interface. The oxide layer on top of the precipitate is clearly bent upwards with the reduction in oxide thickness being negligible. If there is, however, a polysilicon layer on top of the oxide, clear oxide thinning occurs (Fig. 31b). The copper silicide precipitates penetrate into the oxide, and the released silicon interstitials grow epitaxially at the silicon interface in between the precipitates. The arrows in Fig. 31b indicate the silicon regions which lie clearly above the original Si–SiO_2 interface.

3.3.5.3 Stress Fields Induced by Film Edges and Silicon Trenches

The silicon surface is exposed to local stress fields induced by (a) edges of patterned thin films, (b) the oxidation of nonplanar silicon surfaces such as convex and concave trench edges, or (c) filling a trench in the silicon substrate with a material exerting an intrinsic or thermal stress. A review of stress-related problems in silicon technology is given by Hu [38]. Thin films exhibit intrinsic stress due to

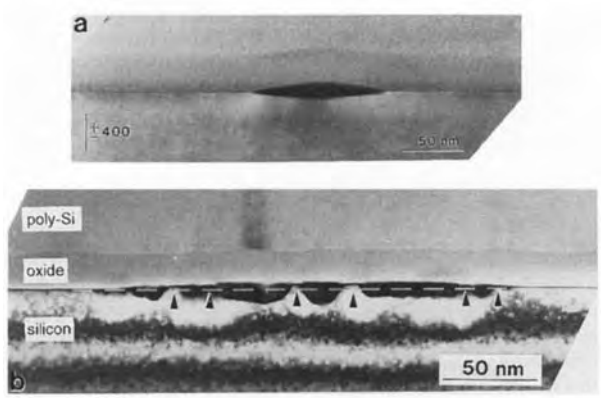


Figure 31. Precipitation of copper-silicide at the (1 0 0)Si-SiO_2 interface. (a) The precipitate bends up the oxide layer. (b) Thinning of the oxide layer by the precipitates in the case of a polysilicon top layer.

their particular microstructure, which is characteristic of the deposition process and the material. Stress in thin films also arises due to the difference in thermal expansion coefficients of the film and substrate. Stress in uniform planar films causes wafer bending. After patterning, however, local stress fields arise at the film edges which may surpass the critical yield stress for dislocation generation. Typical examples are thin Si_3N_4 layers which are used as diffusion barriers for oxygen or as an auxiliary layer acting as a selective etch stop in, for example, contact etching. The local oxidation of silicon (LOCOS) process, which is used to isolate different device areas, uses a nitride layer as a mask, with oxide growth occurring in the mask windows. Due to lateral oxidation below the nitride edge the LOCOS edge assumes the form of a wedge or a 'bird's beak', which causes the nitride mask to bend upwards. If the nitride layer thickness as well as the temperature, time, and ambient atmosphere for oxidation are not chosen carefully, dislocations are generated which glide deeply into the silicon substrate. Detailed TEM work on the dislocation type and their configuration with respect to the nitride edge position is described by Vanhellemont and coworkers [85, 86]. Another example is the dislocation generation in the salicide process (see Sec. 3.3.2 of this Chapter) at the edge of the $TiSi_2$ layer when it exceeds a critical thickness [87].

Silicon trenches are used for device isolation when filled with oxide, or as trench capacitors, and for vertical MOS transistors when covered with a thin dielectric and a highly doped polysilicon layer. Stress fields at concave and convex trench edges develop during thermal oxidation or are generated by the trench filling material due to its intrinsic stress or because of the difference in the thermal expansion coefficients. Typical examples from DRAM technology are given in the literature [76, 79, 88]. The thin dielectric layer at the top edge of capacitor trenches is often accessible to oxygen during various high-temperature processes. This can lead to local oxidation of this area, resulting in a small vertical 'bird's beak'. The additional stress field induced by this wedge may finally cause the nucleation of dislocations. Trench-induced dislocations may be rapidly visualized by defect etching [89]. The local stress fields can be measured by micro-Raman spectroscopy with a resolution of about 1 μm [90]. A higher spatial resolution is offered by simulating the diffraction contrast arising from stress fields in the TEM image or using a fine electron probe ($\leqslant$20 nm) to record convergent beam electron diffraction (CBED) patterns [91]. The CBED methods, however, suffer from the drawback that they are very slow, and at present can barely give timely feedback during technology development.

3.3.5.4 Damage Produced by Ion Implantation and Reactive Ion Etching

Ion bombardment of the silicon surface and subsequent annealing causes residual lattice defects which degrade device performance or act as nucleation sites for secondary defects. Ion implantation for doping of silicon areas operates in the several kiloelectronvolt to megaelectronvolt energy range. The various categories of implantation damage have been described exhaustively [92, 93]. The defects

most frequently encountered are the so-called end-of-range (EOR) dislocation loops, which are of interstitial type. They are formed by the agglomeration of the silicon interstitials, which are generated by displacing silicon atoms from lattice sites and which are driven beyond the projected ion range. The position of these loops with respect to the position of the pn junction is of utmost importance because they may enhance junction leakage. Here we will focus on a special defect which occurs under the mask edge after high-dose implantations of silicon, phosphorus, germanium, or arsenic, which amorphize the silicon substrate above a critical dose [94]. Figure 32a shows a schematic cross-section of a typical MOS transistor with the lightly n-doped drain and source regions (2) extending below the polysilicon gate (cf. Sec. 3.3.4 of this Chapter). The areas denoted by 1 are the high-dose arsenic-implanted source/drain regions. They are completely amorphized after implantation, as can be seen in Fig. 32b at A. The amorphous–crystalline interface smoothly bends to the silicon surface below the spacer oxide mask. Annealing at 500°C for 30 min recrystallizes the amorphous layer epitaxially on the silicon substrate (Fig. 32c). Since the regrowth process is incomplete it becomes obvious that the crystallization front does not follow the original amorphous–crystalline interface but forms a notch under the mask edge. Complete recrystallization as shown in Fig. 32d after 900°C/60 min annealing reveals that a defect has formed at the position of the notch. High-resolution TEM has identified this defect to be a vacancy-type dislocation half loop [94]. Figure 33 gives a schematic explanation of the defect formation. The tangents to

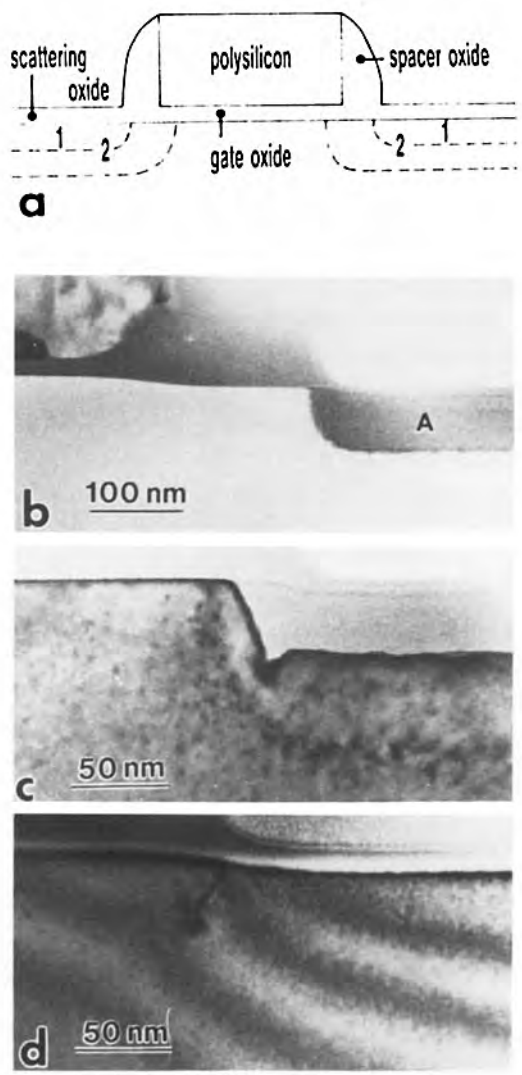


Figure 32. Formation of mask edge implantation defects below the spacer oxide of an MOS transistor due to amorphization of the silicon substrate by a high-dose arsenic implantation. (a) Schematic cross-section of an MOS transistor with source/drain implantations of $5 \times 10^{15}\,cm^{-2}$ arsenic in region 1, and $2 \times 10^{13}\,cm^{-2}$ phosphorus in region 2. (b) After implantation. (c) Partial recrystallization after annealing at 500°C for 30 min. (d) After annealing at 900°C for 60 min.

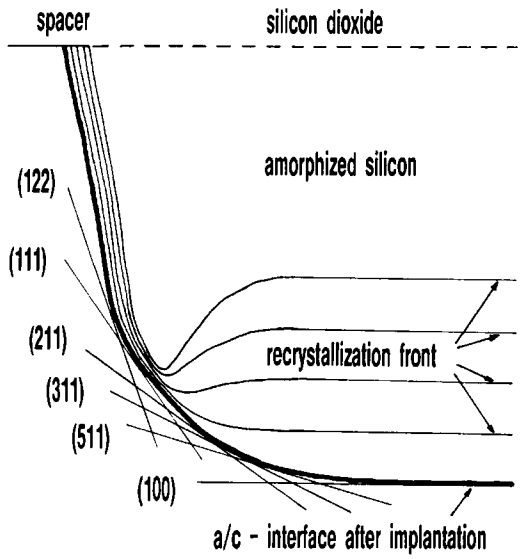


Figure 33. Schematic drawing showing the regrowth at the curved amorphous-crystalline interface which was formed by implantation below the mask edge (cf. Fig. 32b). The recrystallization fronts after isochronal anneals reveal the formation of a notch due to the different regrowth rates on the various lattice planes.

the curved amorphous–crystalline interface below the mask edge visualize the various lattice planes forming this interface. They all have different regrowth rates in the following order: (100) > (110) > (111). The schematic in Fig. 33 shows that a notch must form after isochronal annealings. The mask edge defects have received much attention because high-dose arsenic source/drain implantations are a worldwide standard for n-MOS transistors, and it has been realized [95] that these defects are preferred nucleation sites for secondary defects, such as glide dislocations or metal precipitates, which degrade device performance. It is interesting to note that the amorphizing BF_2 implantation for p-MOS transistors does not produce such defects. The crystallization front produced in the BF_2-implanted area consists of tiny pyramids with {1 1 1} facets. This appearance was attributed to the segregation of fluor at the regrowth front [96]. Thus, the regrowth rate is the same everywhere on this front, and no notch is formed.

Reactive ion etching (RIE) is a dry etching process which allows highly anisotropic pattern transfer with high etch selectivity (see Sec. 3.3.3 of this Chapter). RIE combines physical sputtering with chemically reactive plasma etching. During etching, volatile reaction products are formed by an ion-supported chemical reaction between the attacked material and reactive radicals as well as ions produced in the plasma. Due to the selectivity of the process the etching process stops at the silicon surface once the material above is etched away. However, to ensure complete successful etching, prolonged etching on the bare silicon surface (over-etching) is necessary. Low-energy ions (100–1000 eV) are therefore implanted into the silicon surface during over-etching. A typical patterning step is to open thick oxide layers to form a contact to the silicon substrate. This process requires a CHF_3–O_2 chemistry [97, 98]. An approximately 50 nm thick surface layer is damaged for 600 eV ion energy and 100% over-etching (Fig. 34). The defects consist of platelets on (1 1 1) and (1 0 0) planes being about 20 nm in diameter. High-resolution imaging and Fresnel imaging shows that they contain both silicon atoms and atom species lighter than silicon (carbon and hydrogen from the plasma) [99, 100]. This damage layer has to be removed by an additional dry etching process, for example with NF_3 chemistry, because the RIE-induced defects are nucleation sites for oxidation-induced stacking faults and metal precipitates.

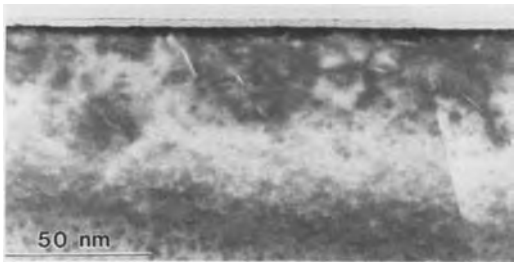


Figure 34. ⟨1 1 0⟩ aligned bright field image of the damage formed by reactive ion etching with CHF_3 near the silicon surface. Platelike defects on {1 1 1} and {1 0 0} planes are visible.

3.4 Silicon Device Structures

3.4.1 In-Line Wafer Assessment by Scanning Electron Microscopy

As pointed out above, fabrication of ICs consists of a complex sequence of all the processing steps described in Sec. 3.3 of this Chapter as well as several others not described such as cleaning steps which call for control of contamination levels. Here, two topics of microscopic characterization will be discussed, that is, application of SEM within the fabrication line and the methods used for failure analysis (see Sec. 3.4.2 of this Chapter). Process control of lithographic and etching processes requires nondestructive assessment within the fabrication line with a short turn-around time. Highly specialized equipment based on optical microscopy and SEM has been developed for this purpose. Optical tools have the advantage of higher throughput. Since the minimum device dimensions are approaching 0.2 μm, the higher resolution of SEM-based tools will further increase their use, and therefore their application will be emphasized here. Detailed contributions to both light and electron optical methods can be found in the literature [101, 102].

In-line wafer assessment by SEM has three main tasks:

- *Inspection*: evaluation of pattern fidelity by, for example, checking for resist and etching residues and monitoring the shape of the sidewalls.
- *Metrology*: measurement of *critical dimensions* (CDs), for example linewidth and pitch (period of regular patterns) as well as placement accuracy, of both resist patterns and etched film structures.
- *Defect review*: First, defects in patterns induced by, for example, particles are detected and localized by automated optical methods. They use either image-processing techniques to compare a certain area with a neighboring area of identical pattern, or detect defects from the scattered light intensity in the scanning laser system. Both methods provide the defect coordinates necessary for detailed characterization (defect review) in the second step. Optical defect review stations are based on, for example, scanning laser microscopes with confocal imaging providing three-dimensional information and allow the operator to classify defect types and perform statistical tracking. SEM-based systems, of course, yield much higher resolution, and thus from the particle shape more easily allow identification of the origin of defects. In addition, EDX spectroscopy can be employed by the operator to obtain elemental analysis.

Special requirements have to be met by scanning electron microscopes for in-line application. Besides clean-room compatibility and oil-free vacuum systems to

minimize contamination of the wafer during imaging, automated wafer-loading systems with cassette-to-cassette operation and wafer stages for 200 mm diameter wafers are required. Of crucial importance is low-voltage operation with electron energies between 0.5 and 2 keV, since standard imaging conditions with high-energy electrons (10–30 keV) lead to charging of the dielectric layers and beam damage of the photoresist patterns, and coating with a conductive film is prohibitive. Recent progress in SEM technology has been summarized by Peters [103]. To obtain good resolution at low beam voltages, field emission (FE) electron sources are required, including either a cold FE cathode or a Schottky emitter, that is, a thermally assisted FE source. Besides the high brightness, FE sources also offer the long service life (up to several thousands of hours) needed for production environment. Also, the lens design has been vastly improved. In virtual immersion lenses (Snorkel lens) the magnetic field extends beyond the lens structure and as such allows a shorter focal length, providing better resolution. In a novel design of a combined magnetic and electrostatic lens, especially low lens aberrations were obtained [104]. This lens yields a resolution of 5 nm even at a 0.6 kV beam voltage. The conical shape of the pole piece allows wafer tilt without loss of resolution due to an increased working distance. In contrast to standard SE detection by their collection at one side of the pole piece, SEs can be extracted through the bore of the final lens and deflected towards a detector. This through-the-lens detection provides symmetric contrast of both sidewalls in CD measurements and allows the bottom of deep holes to be imaged since SEs generated there are also collected. In addition to SEs, back-scattered electrons (BSEs) are also used for imaging since BSEs are less sensitive to charging and can also image the bottom of deep holes [105]. Since in-line scanning electron microscopes play a pivotal role in manufacturing yield improvement, automated operation and wafer throughput are very important. For CD measurements with five sites per 200 mm wafer, a throughput of more than 20 wafers per hour has been achieved.

CD measurements and defect review will be explained in more detail with the aid of two examples each. The photoresist lines in Fig. 35a are 1.5 μm thick and nominally 0.4 μm wide. In the SE image with a 0.6 kV beam voltage they appear as bright regions because SE collection from the trenches between the lines is restricted. The sidewalls are brighter than the center because SE emission is increased at inclined surfaces. Both sidewalls show similar contrast since the CD scanning electron microscope used employs through-the-lens SE detection. For CD measurements the SE signal is recorded across the resist pattern (line scan profiles). Two profiles are shown in Fig. 35a, since the SE detector is split into four sectors, and either the right or the left part was used to record the upper and lower profiles, respectively. Therefore, the profiles are no longer symmetrical, but nevertheless show an increased SE signal at both sidewalls. The splitting of the detector allows automatic distinguishing between holes and elevations. The line scans were not derived from the SE image information but recorded in separate electron scans. This explains the slight lateral shift of the profiles with respect to the SE

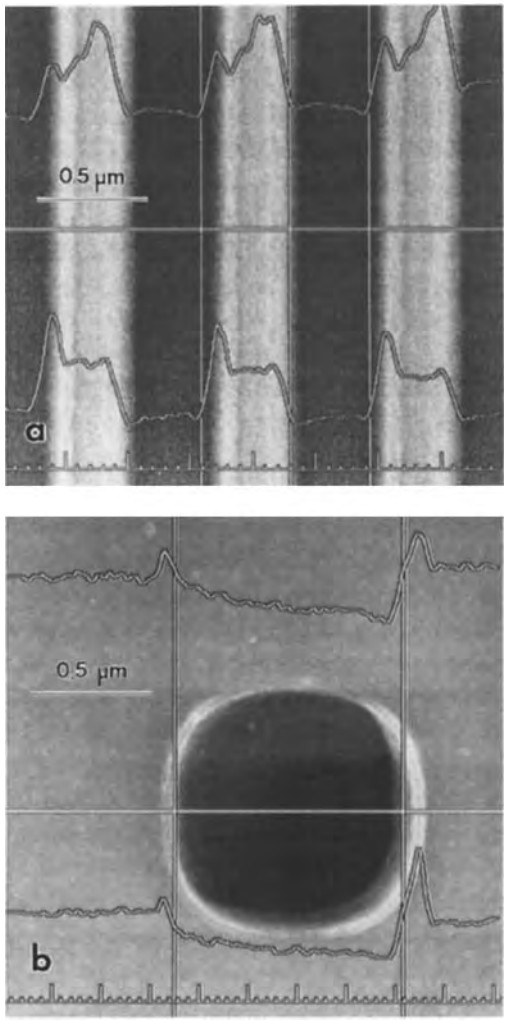


Figure 35. Critical dimension measurements of photoresist lines (a) and contact hole in a BPSG layer (b). Superimposed on the SE images are line scan profiles using the right or left part of a split SE detector.

image. Various algorithms can be applied for automatic SE profile evaluation. Here, the linear regression method was used since the bottom width is of main interest for photoresist lines. Both the baseline and the profile edge are approximated by straight lines, and their intersection defines the position of the bottom edge of the resist. A bottom width of 421 nm was measured in Fig. 35a and a separation of 336 nm between two lines.

Figure 35b shows a contact hole in a BPSG layer of 1.4 µm thickness. With the low beam voltage of 0.8 kV a sharp image of this oxide structure not disturbed by charging effects was obtained. Because of through-the-lens SE collection the bottom of the contact hole can be inspected, and is found to show residual oxide in the upper left part. The sidewalls of the contact hole are not symmetric. Therefore, the increase in SE signal is higher at the right sidewall in both profiles, recorded with the right (upper) or left part of the SE detector (lower profile). For measuring the width of contact holes from the SE profile, different algorithms are applied, for example setting a threshold level for the signal or using the inflexion point. Here, the midpoint between the maximum and minimum of the slope representing the sidewall was chosen. Thus, a value of 0.93 µm was obtained for the width of this hole.

As an example of defect review, Fig. 36a shows a defect found after deposition and patterning of polysilicon lines. The particle in the center of the image has led to an increased polysilicon thickness in its surrounding. After etching, the particle remains, and part of the polysilicon is not removed. This shows that this particle must have originated prior to or at the beginning of polysilicon deposition. A defect embedded in the top layer like this one cannot be classified by optical defect review. In contrast to this defect, the particle of Fig. 36b clearly lies on top of the wafer, with aluminum interconnections forming the top layer. EDX analysis showing chlorine, and the cubic shape of the particles imply that they consist of NaCl.

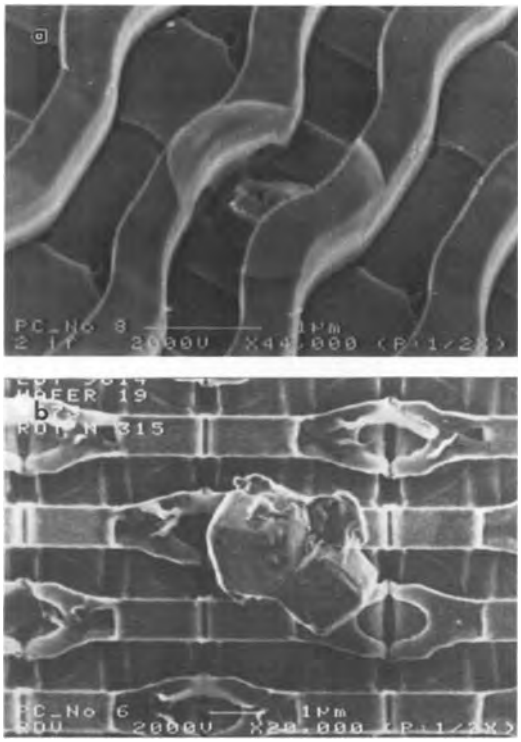


Figure 36. Defect review by scanning electron microscopy. (a) Particle leading to an increased thickness of polysilicon lines. (b) NaCl particles on top of the wafer surface.

3.4.2 Failure Analysis

Due to progress in dimensional miniaturization the physical and logical sophistication of highly integrated circuits increases. Therefore, failure analysis has become more important since precise knowledge of the failure mechanism is necessary first for quick improvement of yield in a fabrication line. Second, improvement of reliability requires clarification of failures after burn-in or failures that occur during operation of ICs. In both cases, failures are traced to incompletely mastered single processes, unreliable, nonoptimized technological parameters, or problems with topography and materials, and so forth. Thus, failure analysis helps to maximize the efficiency of the learning cycle and to transfer the knowledge gained directly into the current fabrication process. A systematic approach for failure analysis after electrical testing first involves localization of the failure site and, second, physical analysis to clarify the failure mechanism. A success rate of more than 90% has been achieved by employing the various techniques systematically for logic ICs with 0.5 μm design rules [106]. Detailed information on all type of techniques for both localization and physical analysis of failures can be found in recent conference proceedings of ISTFA [107] and ESREF [108].

3.4.2.1 Localization of Failure Sites

In special cases the failure site can be derived directly from electrical tests. In refresh analysis of DRAMs, for example, the time after which the first memory cell loses its information is determined. When refresh analysis is performed for a certain block of the cell array, the address of the initially failed cell is obtained. An example of further analysis of the physical cause via TEM characterization of the associated crystal defect is given below. Generally, however, the failure sites have to be located by special techniques depending on the failure type, that is, whether or not the devices show functional failure and/or excessively high total current consumption (I_{cc}) stemming from leakage currents. Functional failures can be located by EBT. I_{cc} leaking produces local heating at the failure site which necessitates, for example, thermal imaging techniques.

EBT is based on the voltage contrast for SE imaging in the scanning electron microscope. When an IC under electrical operation is imaged, SE collection is influenced by the local electric fields prevailing above the surface of the IC. Thus, interconnections with negative voltage appear bright since SE collection is facilitated by these fields. Interconnections with positive voltage on the other hand appear dark: part of the emitted SEs cannot be collected because the local electric fields deflect them back to the surface. Apart from voltage contrast imaging, voltages can be measured locally with an electron spectrometer for SE collection. Furthermore, stroboscopic techniques allow waveform measurements of electric signals on internal nodes in a circuit and mapping of logic states as function of time. An introduction to EBT is given by Fuchs et al. [5], an overview on application by Wolfgang [109], and more detailed information can be found in reference [110]. When EBT is applied to failure localization, major faults often can be identified by comparing static voltage contrast images of the nonfunctional device to those of a functional device. Alternatively, internal electrical signals have to be traced by waveform measurements until the critical fault site is located. In addition to failure analysis, EBT is also employed for design verification of ICs, providing localization and quantification of design weaknesses.

The source points of leakage currents on an IC can be located by thermal imaging (hot spot detection) since electrical energy is converted into heat at these points. In many cases, hot spots indicate technology defects such as short circuits in aluminum or polysilicon levels, substrate leaks around contact holes, gate oxide breakthroughs, and floating gates. They may also indicate design weaknesses such as increased current paths, faults in the layout, or latch-up regions in CMOS circuits [111]. A recent overview on thermal imaging techniques is given by Kölzer et al. [112]. In liquid crystal thermography (LCT), which is widely used for hot spot detection, a thin layer of a liquid crystal (LC) is applied onto the surface of an IC [113]. The device is then examined under operation in an optical microscope using polarized light with the polarizer and analyzer crossed. Areas without the LC appear dark. The optical properties of the LC depend on the temperature. Below the clearing-point temperature T_c the LC exhibits crystalline order since nematic LCs are used as a rule. Birefringence in the LC causes the polarization plane of the light to rotate, resulting in some brightening of the image. On IC areas with a local temperature above T_c, the LC transforms into the optically isotrope phase. Here, the polarization plane remains unchanged and the relevant area appears dark (hot spot). By controlling the temperature of the device with a heating stage to be just below T_c, hot spots with low power dissipation can also be detected. In practical operation, temperature differences of approximately 0.1°C can be resolved. Depending, for example, on their depth below the surface of the IC, failures with a power dissipation of 50–500 μW can then be detected [114]. The spatial resolution of the method is about 1 μm. Not every detectable hot spot represents a failure site, since specific parts of an IC may have an increased temperature during standard operation.

LCT allows quick detection of hot spots but provides only a binary response, indi-

cating whether or not the hot area is above the transition temperature. In fluorescent microthermography (FMT), on the other hand, the temperature behavior of a device can be studied with improved temperature resolution (<0.01°C) and spatial resolution (0.3 μm) [115, 116]. FMT involves applying a thin film of europium thenoyltrifluoroacetonate (EuTTA) and polymethyl methacrylate (PMMA) dissolved in acetone onto the surface of the IC. Illumination with ultraviolet light in a wavelength band of 340–380 nm stimulates fluorescence in the film at 612 nm. No significant absorption occurs in the film above a wavelength of 500 nm causing secondary fluorescence. Thus, a clear separation between excitation source and fluorescent light is possible. The quantum efficiency of the 612 nm fluorescence decreases exponentially with temperature, thereby providing information on the temperature distribution of the operating IC. The fluorescence radiation is acquired by a charge-coupled device camera.

In addition to local heating, the excessive currents at failure sites may lead to the emission of light. By combining optical microscopy with the superb amplification of night-vision technology, very faint light emission in the visible and near-infrared region is detected, and the location of emission pinpointed. An overview of emission microscopy is given by Kölzer et al. [117]. Light emission is induced mainly by three phenomena:

- Recombination of minority carriers injected into the pn junction, which enables the analysis of forward-based bipolar devices as well as parasitic bipolar effects such as latch-up phenomena in CMOS devices.
- Field-accelerated carriers, which are generated by high local electric fields, allow the investigation of, for example, saturated MOS transistors, hot electron degradation, effects of electrostatic discharge events, or of leakage currents due to microplasma discharge at dielectric breakdown.
- Tunneling (Fowler–Nordheim) currents flowing through thin dielectrics offer the potential to study gate oxide quality.

Emission microscopy goes beyond the mere localization of failures since the emission spectrum provides information on the defect mechanism [117]. Since light emission is easily detected if the leakage current flowing at the source point is more than 10 μA, emission microscopy is the most useful technique for quick localization of leakage current sources in the substrate or related to thin dielectrics [106]. Of course, the application of emission microscopy is hindered when large areas of metallization cover the failure site. Failures within the various metallization levels are located by hot spot detection.

3.4.2.2 Physical Failure Analysis

To identify the cause of a failure after its localization, detailed physical analysis is required involving various procedures of destructive specimen preparation. Usually, the first step is 'delayering', that is, selective removal of technology layers down to the level of failure by wet chemical etching, RIE, or mechanical polishing. In many cases the origin of the failure can be clarified by SEM inspection. If this is not possible, cross-sections provide additional information. Precision cross-sectioning, that is, preparation of a cross-section

through the preselected area of the failure site, is achieved by well-controlled polishing for SEM and TEM samples, but is also possible without polishing by focussed ion beam (FIB) techniques.

For polishing of SEM cross-sections the sample is not embedded but mounted on a fixture allowing observation of both the surface and the cross-section of the IC in an optical microscope [118]. Thus, the polishing process can be controlled to an accuracy of less than 0.2 μm for the position of the section plane. The cross-sections can be imaged without further preparation when low beam voltages of a few kilovolts are used in the scanning electron microscope. Alternatively, slight etching delineates the various layers more clearly by creating contrast-enhancing steps at the interfaces between the layers. The etching solutions used either attack SiO_2 or silicon [119, 120]. Sputter coating with a conductive gold–palladium film facilitates imaging of dielectrics but slightly degrades the resolution.

A failure causing high standby currents in a 16 Mbit DRAM was located by LCT and analyzed in detail [121]. In Fig. 37a the operator marked the area by a circle where the dark 'hot spot' occurred first when the IC temperature was increased. No indication of any defect could be found on the chip surface (Fig. 37b). The technology layers were successively removed by face lapping with the polishing procedure being controlled by microscopic inspection. Only when the bit line material was removed was a defect covered by an oxide layer detected. After slight etching of the oxide a protrusion is clearly seen in the SEM image (Fig. 37c, arrow). A precision cross-section allows identification of the failure origin (Fig. 37d). The protrusion was caused by a local increase in polysilicon thickness. It is well known that particles can cause such defects during polysilicon deposition. Since the region of increased polysilicon thickness was not removed during pattering of this level, the polysilicon residue formed a connection to the silicon substrate and thus led to an increased standby current.

When crystal defects are expected to cause the failure, TEM cross-sections or plan view specimens may be used for their identification. This applies to defects reducing the retention time of memory cells during refresh analysis of DRAMs. Defect-induced leakage currents through the substrate diminish the charge stored on the substrate node of the trench capacitor. For preparation of a precision cross-section, the defective trench capacitor has to be marked by, for example, laser shots or FIB in the surroundings. Polishing from one side is analogous to the procedure for SEM cross-sectioning using the laser shots to identify the defective trench. After polishing from the other side, final ion milling has to be controlled to obtain electron transparency at the selected site. In the example of Fig. 38a small precipitates were indeed found in the vicinity of a trench capacitor with reduced retention time. Figure 38b shows that these defects are platelets on {1 1 1} planes with a diameter of up to 40 nm exhibiting strong strain contrast [122]. After additional ion milling a ⟨1 1 0⟩ lattice image of the same precipitate could be obtained (Fig. 38c) showing two additional {1 1 1} planes inserted into the silicon matrix. In EDX analysis a small copper peak was found (Fig. 38d), providing evidence for copper precipitates stemming from metal contamination during device processing.

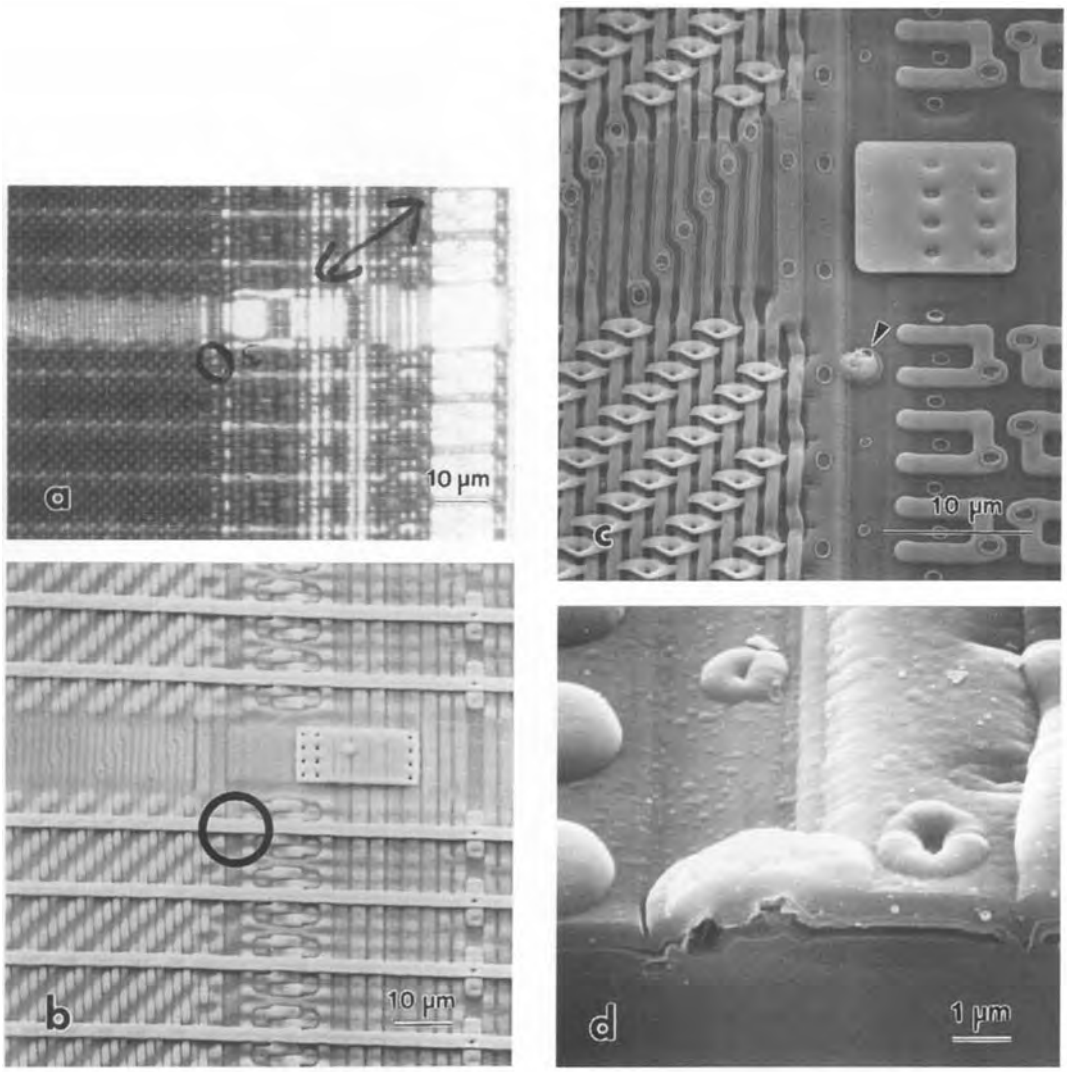


Figure 37. Failure analysis of a 16 Mbit DRAM with a high standby current. (a) Optical micrograph with the 'hot spot' marked by a circle after detection by LCT. (b, c) SEM images of the same area as in (a). Only after removal of several technology layers is the defect visible (c). (d) Precision cross-section through the defect.

FIB systems have advanced to a stage where they are now routinely applied for both circuit modification and failure analysis. Liquid metal ion sources employing Ga^+ ions with an energy of 20–30 keV provide beam diameters well below 50 nm and high current densities ($>10\,A\,cm^{-2}$) [123]. Analogous to SEM, the ion beam is scanned across the specimen surface for imaging and the SEs released by the Ga^+ ions serve as the image signal. Beams with a diameter below 10 nm and a current of only ~1 pA are employed for imaging presently, and change the sample only slightly. When much higher currents (several nanoamps) are used, local sputtering with high rates

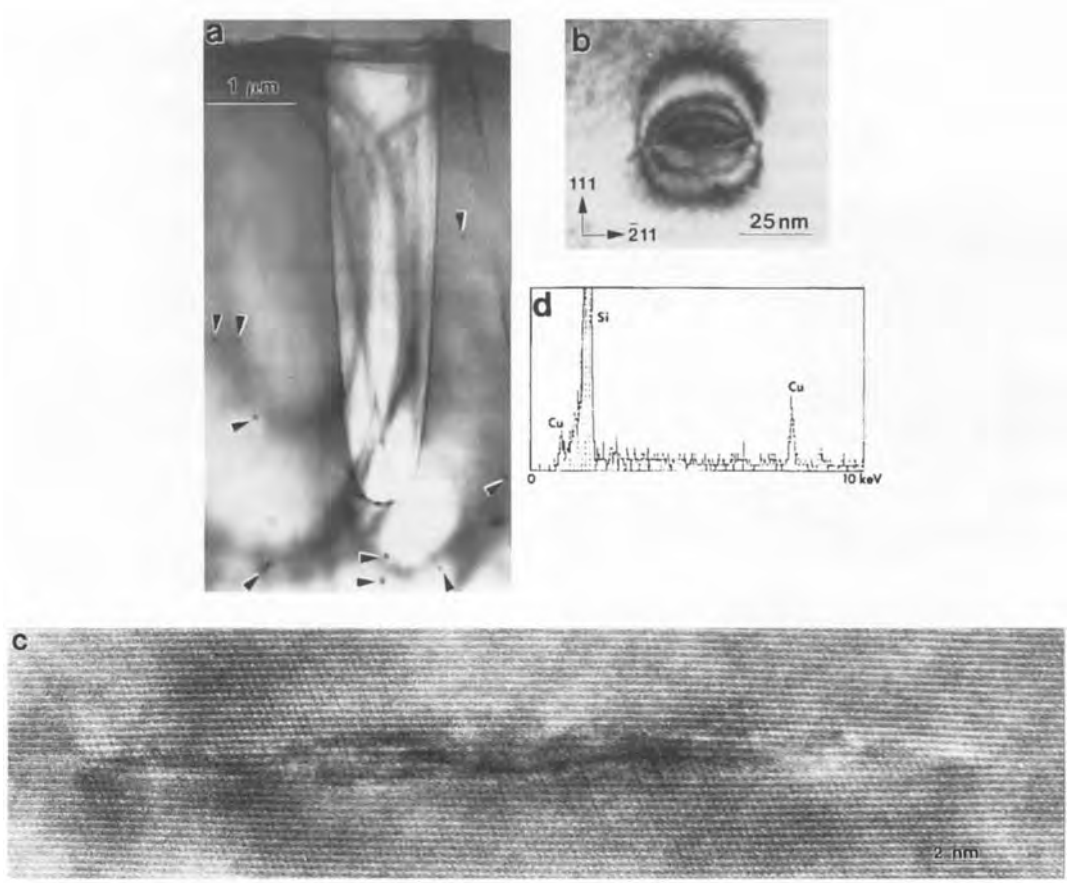
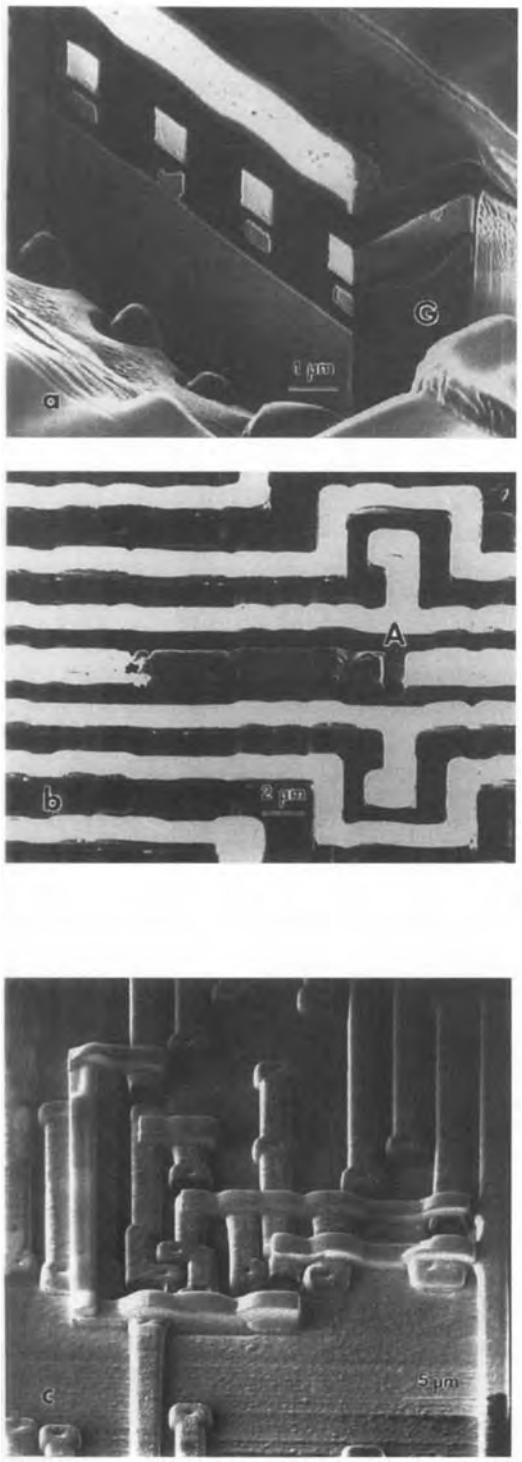


Figure 38. TEM failure analysis of a specific defective storage cell in a 4 Mbit DRAM device. (a) Bright field image of the defective storage capacitor. The arrows mark precipitates. (b) Bright field image of a single precipitate. (c) ⟨1 1 0⟩ HREM image of a precipitate having two additional {1 1 1} planes. (d) EDX spectrum of the precipitate showing copper.

yields the possibility of micromachining, such as for etching a cross-section at a location selected from the image. This combination of micromachining and high-resolution imaging allows very quick physical failure analysis, and FIB workstations are therefore increasingly being used for this purpose [124].

In the FIB image of Fig. 39a, cross-sections along two perpendicular planes were etched into an IC. The narrow section cuts through a polysilicon gate extending over the field oxide on both sides of the gate region (G). The fault in this device was located at the gate in the middle of the longer section plane. It is clearly seen that the field oxide below the polysilicon is missing here, probably because of a mask defect. The microscopic cross-sections prepared by FIB can also be imaged in the SEM when superior resolution is required (see Fig. 1).

FIB image contrast is different from SE images recorded by SEM. Since the SE yield, that is, the number of SEs emitted per incident primary ion, is larger than



unity, dielectrics such as oxides always charge up positively and therefore appear dark. Positive charging restricts the SE collection analogously to the voltage contrast used in EBT. This charging, however, usually does not degrade image quality since the positively charged areas can attract SEs emitted from other parts of the specimen, and the much higher mass of the Ga^+ ions compared to electrons makes them less sensitive to local electric fields. The SE signal generated by primary ions is very surface sensitive since the ions penetrate into the specimen only a few tens of nanometers and the emitted SEs have low energy (<10 eV) [123]. Furthermore, the SE signal exhibits no 'material contrast' as in SEM because the fraction of backscattered electrons contributing to the SE signal in the SEM is missing in FIB systems. Since the penetration of the primary ions depends sensitively on crystal orientation (channeling), the SE contrast also varies. This effect allows imaging of the grain structure of metallization layers, such as for aluminum interconnections. In Fig. 39b charging was used intentionally to locate the interruption of an aluminum interconnection. After cutting the aluminum line (at A in Fig. 39b), the part between the cut and the interruption is disconnected from the next contact and therefore charges up positively during ion bombardment. This part thus appears dark and allows easy location of the failure.

The capability of FIB for micromachining also allows the preparation of precision

Figure 39. Focused ion beam application in failure analysis. (a) FIB image of precision cross-section etched by FIB. (b) FIB image of aluminum interconnections. The interruption is visible because of local charging. (c) Ion-beam-induced deposition of metal interconnections (horizontal lines).

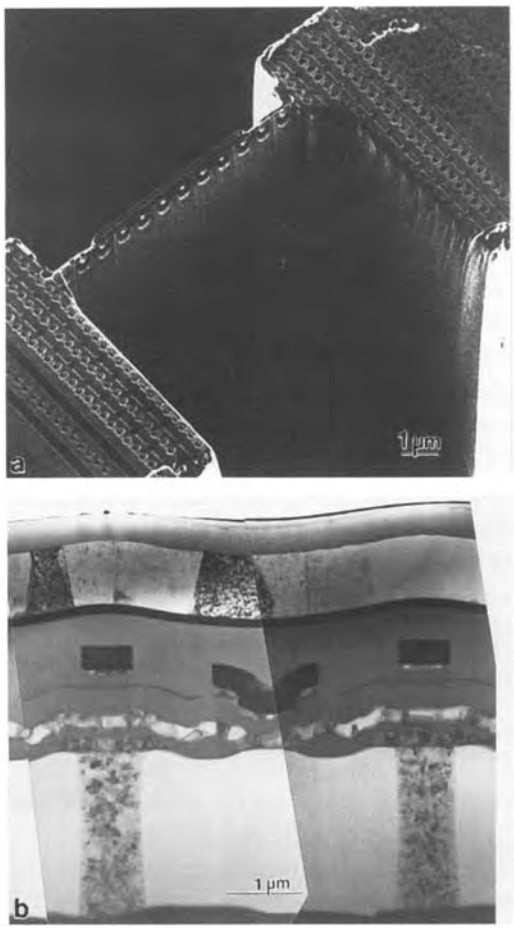


Figure 40. FIB preparation of precision TEM cross-sections. (a) FIB image of a thin lamella through the storage cell array of a DRAM prepared by FIB. (b) TEM image obtained from the thin lamella.

cross-sections for TEM inspection [125]. The thin slice in the FIB image of Fig. 40a represents the electron transparent part of the sample. To protect the surface, a narrow tungsten line was deposited (see below) prior to FIB etching. The TEM image of this specimen (Fig. 40b) shows the section plane to cut through the trenches of the 4 Mbit DRAM.

Besides failure analysis, FIB systems are widely applied for circuit modification using two operations: cutting of unwanted interconnections and depositing new ones. This possibility allows designers to test modified circuits with short cycle times and without the need to generate new mask sets and silicon. An example of cutting was shown in Fig. 39b. Redeposition during etching of deep holes with high aspect ratios can be avoided by gas-assisted etching. Reactive gases such as halogens volatilize the sputtered material and enhance the FIB milling rate [126]. For deposition of interconnections, organometallic gases are injected near the area of interest. During ion beam-induced deposition, gas molecules adsorbed at the surface are decomposed by the ion beam with the metal remaining in the areas scanned by the ion beam. All the horizontal platinum interconnections in Fig. 39c were deposited in a FIB system.

3.5 Compound Semiconductor Devices

3.5.1 Heteroepitaxial Layer Structures

Modern compound semiconductor devices employ heteroepitaxial layer structures which are grown by metal–organic vapor phase epitaxy (MOVPE) or MBE [127]. Both electronic and optoelectronic devices make use of single and multiple quantum well structures with layer thicknesses being typically in the range of several nanometers. The 2D electron gas confinement achieved with $Al_xGa_{1-x}As/GaAs$ and $Al_xGa_{1-x}As/In_xGa_{1-x}As/GaAs$ layers finds application in high electron mobility transistors (HEMTs, see Sec. 3.5.2 of this Chapter).

Binary (AB), ternary ($A_xB_{1-x}C$), and quaternary ($A_xB_{1-x}C_yD_{1-y}$) III–V compounds with the sphalerite structure are used with high success for layer structures in optoelectronic devices (e.g., light-emitting diodes and lasers), because the band gap varies with composition. However, in most cases a change in composition causes the lattice constant to change too. Thus, band gap engineering not only requires adjustment of the width of the band gap by composition but also exact matching of the lattice constant of the layer to that of the GaAs or InP substrate. For optimum device characteristics layer growth has to be controlled precisely with respect to thickness and uniformity of the layers, abruptness of the hetero-interfaces, composition and lattice match, and layer doping.

Analysis of heteroepitaxial semiconductor layers first of all involves X-ray diffraction to examine whether the correct lattice match was obtained in the case of unstrained layers or to determine the lattice distortions in the case of strained layers. Multiple crystal diffractometers provide the high angular resolution required for this purpose. Accurate determination of layer thicknesses and compositions is achieved by simulation of the diffraction profiles using the dynamic theory of diffraction [128]. This procedure is usually more straightforward for superlattice systems than for single layers or aperiodic layer structures. Also, photoluminescence (PL) is frequently used to assess heterostructure layer systems [127], but in the case of quantum wells often relies on accurate layer thickness determination by other methods in order to separate, for example, effects of layer composition and thickness on the spectral position of the PL lines. Both techniques have in common that they usually measure averages over an area of some hundred micrometers across, and/or they form the average over several periods in a multiple quantum well structure. The depth profile of dopants can be measured by SIMS, as in silicon technology [5, 60].

Cross-sectional TEM imaging is currently the only technique which offers valuable information with high spatial resolution both in the vertical and lateral directions. The points of interest for TEM analysis comprise determination of layer thicknesses (even below 1 nm), interface abruptness, layer composition, and uniformity of the layers. The assessment of the interface abruptness requires chemical and/or structural information down to the atomic level. This is the case for compositionally sharp or graded interfaces and for the detection of interface steps. Moreover, imperfect growth conditions stemming from too high lattice mismatch, surface steps, or growth disturbances as well as relaxation mechanisms in strained layer systems may give rise to the formation of crystal defects such as dislocations, stacking faults, antiphase domains, and precipitates. Microscopy of lattice defects is not discussed further here, and selective reference is made to Mahajan [129] and Ueda [130]. Basically, three approaches exist to characterize heteroepitaxial layer structures: (a) the {200} dark field imaging of thin cross-sections, (b) imaging of cleaved wedges, and (c) high-resolution imaging.

3.5.1.1 {200} Dark Field Imaging of Thin Cross-sections

The preparation of thin cross-sections for TEM has become a routine method within

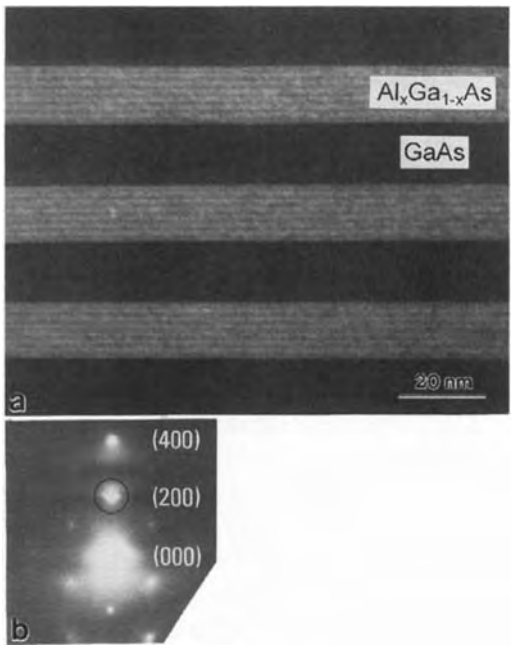


Figure 41. (a) {200} dark field image of an $Al_xGa_{1-x}As$/GaAs multi-quantum-well structure. The bright and dark stripes in the $Al_xGa_{1-x}As$ layers reveal a fluctuation in the aluminum content. (b) Diffraction pattern. The (200) reflection used for the dark field image (a) is encircled.

the last decade, and a description may be found for example in Cerva and Oppolzer [131]. The cross-sectional specimen has to be oriented to the electron beam such that the layer interfaces are viewed end-on. A {200} dark field image is obtained by allowing only the {200} reflection to pass through the objective aperture, as indicated in Fig. 41b. The resolution of this technique of approximately 0.5 nm is limited by the aperture cutting off electrons with larger scattering angles. Also, the apparent interface width increases with specimen thickness because of tilting to the Bragg angle necessary for exciting the {200} reflection. In the {200} dark field image of Fig. 41a the $Al_xGa_{1-x}As$ barriers and GaAs wells of a multiple quantum well structure appear bright and dark, respectively. This may be understood by considering that the kinematic intensity is proportional to the square of the structure factor F, which for the {200} reflection of a binary compound having the sphalerite structure is proportional to the difference between the atomic scattering amplitudes f of the cation and anion. The structure factor F_{200} for GaAs is almost zero because $f_{Ga} \approx f_{As}$, whereas in the case of AlAs there is a large difference. For the ternary compound $Al_xGa_{1-x}As$ the structure factor is given by $F_{200} \approx 4x(f_{Al} - f_{Ga})$. The contrast between layers of different composition is best described by the (200) kinematic intensity ratio $R(x)$ of a ternary and a binary compound ($A_xB_{1-x}C$ and BC, respectively). $R(x)$ is given by

$$R(x) = I(A_xB_{1-x}C)/I(BC)$$
$$= \{1 + [x(f_A - f_B)/(f_B - f_C)]\}^2$$

yielding

$$R(x) = (1 + 2.91x^2)$$

for $Al_xGa_{1-x}As$/GaAs at 200 kV. $R(x)$ increases monotonically with increasing aluminum content x for the system $Al_xGa_{1-x}As$/GaAs. Thus, $Al_xGa_{1-x}As$ layers always appear brighter than GaAs layers, and a qualitative interpretation of the {200} dark field image contrast is possible. For example, the $Al_xGa_{1-x}As$ barriers in Fig. 41a show fine alternating bright and dark fringes in the $Al_xGa_{1-x}As$ barriers. This modulation in the aluminum content is attributed to small variations in the flux profile of the group III elements over the substrate surface [131]. Differences in the aluminum content of $\Delta x > 0.05$ can be recognized from such images [132], and by quantification of the

contrast using densitometry the aluminum content x may be determined with an accuracy of about 3% [133]. The $In_xGa_{1-x}As/GaAs$ system is an example where the contrast does not change unambiguously with the indium content x [134], and hence compositional variations have to be identified by analytical TEM. In general, imaging of conventional cross-sections has the advantage that the layers may be inspected along a relatively large distance (a few tens of micrometers). This is especially useful when assessing wavy interfaces, layer inhomogeneities, and crystal defects. Further examples of the application of this technique may be found in Cerva and Oppolzer [131].

3.5.1.2 Imaging of Cleaved Wedge Specimens

Conventional TEM cross-sections are particularly useful for studying laterally nonuniform effects. In most instances, however, only depth information is required, which can also be obtained by imaging the edge of a cleaved wedge specimen [135]. Since specimen preparation by cleavage is much faster than the conventional procedure, this technique has become very popular. It exploits the property of compound semiconductors such as GaAs and InP to cleave precisely along {1 1 0} planes. A wedge-shaped specimen with a sharp 90° edge (Fig. 42a) can be produced by cleaving a [0 0 1]-oriented wafer along two {1 1 0} planes which are normal to each other. This preparation procedure takes approximately 10 min, and the specimen can then be transferred immediately into the microscope. If the cleaved edge is viewed in transmission after orienting it perpendicular to the electron beam, a cross-sectional image of the heterostructure is obtained.

Just as with thin cross-sections, the wedge-shaped specimens (Fig. 42b) can be imaged with the {2 0 0} dark field method when the sample is tilted slightly away from the [1 0 0] zone axis. For the dark field image, however, that type of {2 0 0} reflection has to be selected which has its diffraction vector parallel to the edge of the sample. Otherwise the layer interfaces are not viewed in projection. In Fig. 42b, the cleaved edge is on the left of the image (arrowed) and is not directly visible. The thickness increases towards the right in accordance with the geometry of the specimen (Fig. 42d). In the thin regions immediately beside the cleaved edge, the intensity in the $Al_xGa_{1-x}As$ layers increases with thickness. Thickness fringes, however, do not appear in thicker regions because the extinction distance of the {2 0 0} reflection is very large. As the result of increased electron scattering and 'absorption' by the objective aperture, intensity drops again at distances of more than approximately 60 nm from the cleaved edge.

In addition to displaying the layer structure, another imaging method allows compositional analysis [135]. This entails tilting the wedge specimen exactly into the [1 0 0] zone axis and recording a bright-field image, that is, the small objective aperture allows only the directly transmitted beam to pass through. The precise orientation is again controlled with the aid of the diffraction pattern. Multiple beam conditions result, because numerous reflections are excited. Figure 42c shows such a bright field image of the $Al_xGa_{1-x}As/GaAs$ layer structure of

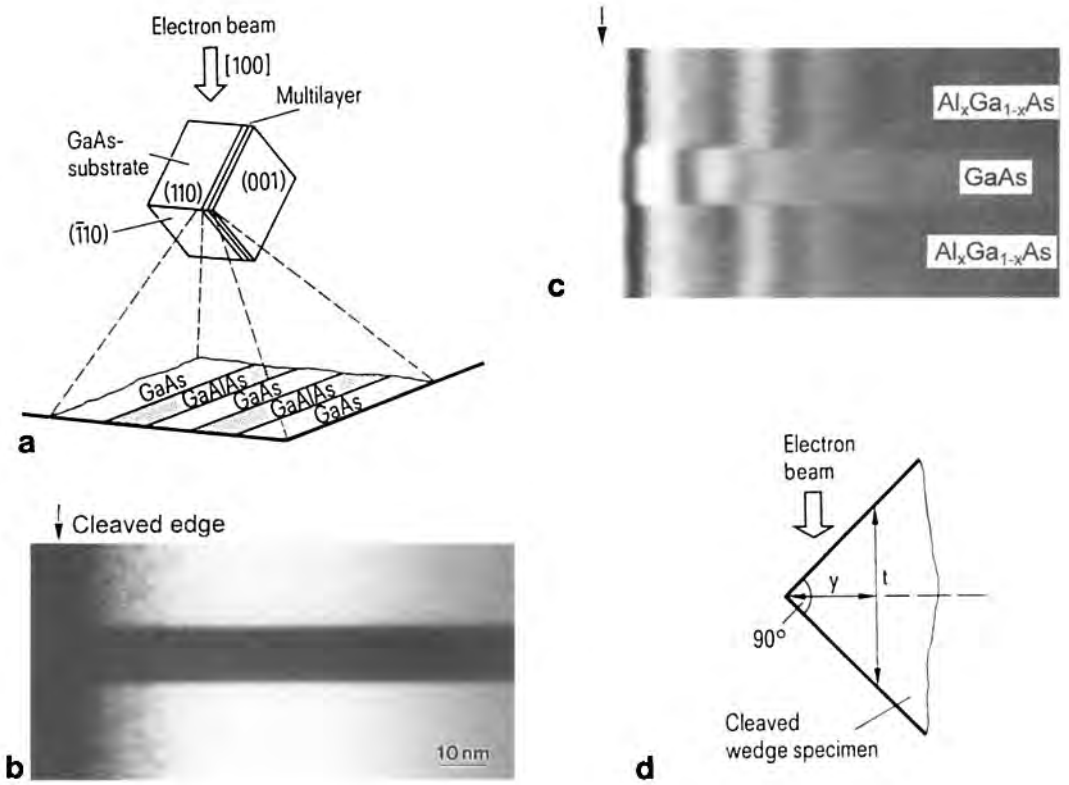


Figure 42. TEM analysis of cleaved wedge specimens. (a) Schematic showing the geometry of the specimen with respect to the electron beam and the cross-sectional image of the multilayer. (b) {2 0 0} dark field image of an $Al_xGa_{1-x}As/GaAs$ quantum well structure. (c) ⟨1 0 0⟩ aligned bright field image of the same quantum well structure as in (b). (d) Geometric relationship between transmitted sample thickness t and distance y from the cleaved edge.

Fig. 42b. Thickness fringes parallel to the cleaved edge can now be seen in the layers. The defined geometry of the 90° wedge means that the position of the thickness fringes (distance y from the cleaved edge) correlates unambiguously with specimen thickness t: $y = t/2$ (Fig. 42d). The periodicity of the depth oscillations and hence the position of the thickness fringes depends on the structure factor, which in turn varies with the aluminum content of the layer. The fringes in layers of different aluminum content are thus displaced with respect to each other. When the position of the thickness fringes is calculated with the aid of the dynamic theory of electron diffraction, the aluminum content can be quantitatively determined [135, 136]. As with the {2 0 0} dark field images, the abruptness of the interfaces can be assessed with a resolution of about 0.5 nm. Bending of the thickness fringes within a layer or at an interface is indicative of a change in the aluminum content or a transition region at the interface.

The two $Al_xGa_{1-x}As/GaAs$ multiple quantum well structures in Fig. 43 were grown by MOVPE within the same

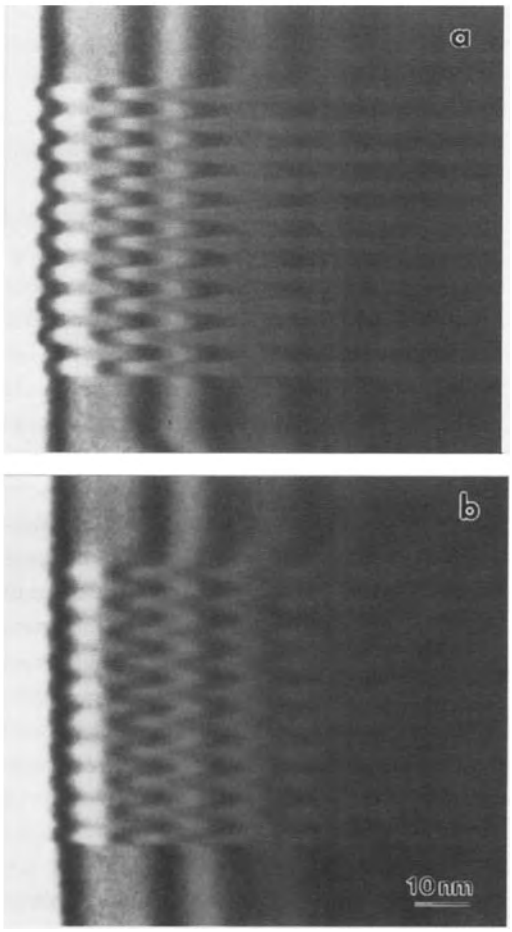


Figure 43. ⟨1 0 0⟩ aligned bright field images of cleaved wedge specimens. The $Al_xGa_{1-x}As/GaAs$ multi-quantum-well structure was grown by MOVPE with growth interruptions at the interfaces (a), and without interruptions (b).

specimen so that they could be imaged by TEM at the same time and under the same conditions. The multilayer in 'Fig. 43a was deposited with growth interruptions at the interfaces, and exhibits abrupt interfaces. In contrast, the multilayer in Fig. 43b exhibits transition zones at the interfaces approximately 1.5 nm in width, because growth was not interrupted. A variety of further examples making use of the cleaved wedge preparation can be found in the literature [131, 134, 137].

3.5.1.3 High-Resolution Imaging

HREM can be applied to both thin cross-sections and wedge specimens, and provides not only structural but also quantitative chemical information on layer interfaces at atomic or near-atomic scale. This requires the HREM image intensities in the dot pattern of the lattice to be related to the chemical composition. Then, by comparing the image intensities in the specifically different contrast patterns at the interface and on both sides of the interface, the chemical variation across the interface can be derived. An algorithm was developed to extract the chemical signal quantitatively under suitable imaging conditions [138]. Over the last few years this real-space algorithm has been extensively used on images projected in the ⟨1 0 0⟩ orientation for various systems such as $Al_xGa_{1-x}As/GaAs$ and $Hg_xCd_{1-x}Te/CdTe$ to study structural and chemical interface properties and point defects. A theoretical and experimental treatment of the image contrast in ⟨1 0 0⟩ projection based on the nonlinear imaging theory was described for the $Al_xGa_{1-x}As/GaAs$ system [139, 140] and silicon/germanium system [141]. The algorithm described by Thoma and Cerva [140] was used to verify the different abruptness of AlAs/GaAs and the GaAs/AlAs interfaces [142], and an example is given below. Recently, a general approach to the measurement of the projected potential in crystalline solids in any projection without knowledge of the imaging conditions has been given [143, 144].

As an example, the HREM assessment of the interfaces in an AlAs/GaAs double-barrier quantum well (DBQW) structure will be described in more detail. The interface quality of DBQWs used in resonant tunneling structures determines their electrical properties and, therefore, the interface quality is of decisive importance. Figure 44a shows a ⟨1 0 0⟩ HREM image of a conventional thin cross-section from an AlAs/GaAs DBQW structure grown by MBE [142]. In order to analyze the interface abruptness quantitatively with the algorithm described by Thoma and Cerva [140], the specimen thickness and focus have to be within the optimum range ($7\,\text{nm} < t < 13\,\text{nm}$ and $\Delta f = -25 \pm 5\,\text{nm}$ at 400 kV) so that a clear difference in contrast between the GaAs and AlAs layers appears, just as in Fig. 44a. The AlAs layer is characterized by a strong (2 0 0) pattern contrast, with the white dots corresponding to aluminum atom columns and black dots to arsenic atom columns, whereas in GaAs a (2 2 0) pattern contrast prevails, where white dots represent the positions of gallium or arsenic atom columns and black dots are tunnels in the structure. When crossing the interface, one type of contrast pattern changes into the other gradually. Under the given imaging conditions, this change in contrast with the aluminum content can be described by those Fourier components of the image intensity which contain the linear beam interactions between the transmitted and the (2 0 0) beams as well as the nonlinear interactions between (2 0 0) and (2 2 0) beams [139, 140]. This behavior stems from the fact that in the ⟨1 0 0⟩ projection the amplitudes of the (2 0 0) beams in $Al_xGa_{1-x}As$ increase with the aluminum content x whereas those of the (2 2 0) beams remain almost constant for the chosen imaging conditions. For determination of the local composition the contrast pattern is divided into unit cells which contain all the chemical information for the quantification procedure. In the case of $Al_xGa_{1-x}As$ in the ⟨1 0 0⟩ projection, the smallest cell has a size of $0.28 \times 0.28\,\text{nm}^2$ and is square shaped [138, 140] because the chemical

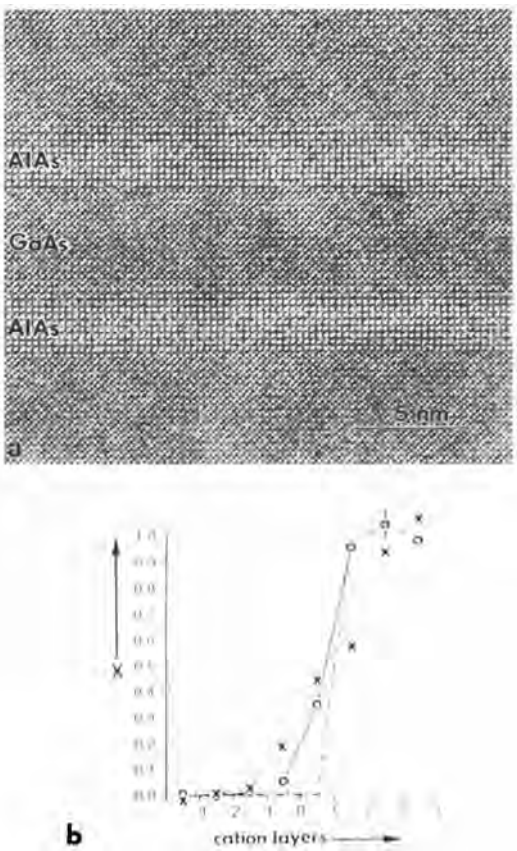


Figure 44. (a) ⟨1 0 0⟩ HREM image of an AlAs/GaAs double barrier quantum well structure grown by MBE showing characteristic dot patterns for AlAs and GaAs. (b) Aluminum concentration profiles across the interfaces obtained by quantitative evaluation of the dot pattern contrast. Sharp normal (AlAs on GaAs) interface (⊙), diffuse inverted (GaAs on AlAs) interface (×), and ideal interface (· – · – ·).

information is contained in the (200) beams which have diffraction vectors corresponding to real-space distances of 0.28 nm. In the $Al_xGa_{1-x}As/GaAs$ system the experimental accuracy of the method to determine the aluminum content in one individual cell of the contrast pattern was shown to be ± 0.1 [140]. Averaging over many cells along an interface improves this value drastically. Specimen preparation is an important issue for quantitative HREM imaging because it determines the signal-to-noise ratio in the image. Our laboratory experience confirms that the highest signal-to-noise ratio is found in chemically etched $Al_xGa_{1-x}As/GaAs$ samples, and the lowest in argon ion beam-thinned specimens [138]. The quality of cleaved wedge samples lies inbetween, and benefits from the simple preparation procedure. Layers with very low aluminum content ($x \approx 0.3$) are better imaged in chemically etched samples.

The clear difference in interface abruptness between normal (AlAs on GaAs) and inverted interfaces (GaAs on AlAs) can already been seen by visual inspection of the HREM image in Fig. 44a: the normal interfaces appear abrupt and the inverted ones diffuse. To quantify this impression, the composition profiles across the interfaces were determined using the algorithm described above (Fig. 44b). By averaging over 64 cells along the individual interfaces, the statistical error for the aluminum content x in the atomic layers of the transition region was reduced to $\Delta x = \pm 0.04$. The normal interface exhibits a transition region extending over one atomic layer, whereas the diffuse inverted interface extends over three layers. This difference between normal and inverted interfaces can be explained by the higher surface mobility of gallium atoms on the GaAs surface compared to that of the aluminum atoms on the $Al_xGa_{1-x}As$ surface during MBE growth. These results are in perfect agreement with electrical transport measurements [142].

A drastic improvement of HREM imaging in the ⟨1 1 0⟩ projection was obtained by removing the ion milling damage by chemical etching. Then, $Al_xGa_{1-x}As/GaAs$ and SiGe short-period superlattices can be imaged under certain conditions with high contrast [145], revealing interface steps and ordering effects clearly. This method was recently compared to the quantitative imaging method in the [1 0 0] projection by using the same samples and was found to yield the same results [146].

3.5.1.4 High-Resolution Chemical Analysis

High-resolution Z-contrast imaging combines very high compositional sensitivity with atomic resolution and has been applied to various semiconductor interface problems [147]. A high-current electron beam with a diameter of 0.2 nm or less generated by a field emission gun is scanned across a cross-sectional sample oriented in, for example, the ⟨1 1 0⟩ or ⟨1 0 0⟩ direction. The electrons scattered incoherently to high angles are collected below the specimen by an angular dark-field detector. This allows HREM images to be produced with strong differences in image intensity for elements with different atomic number Z, because the high-angle scattering cross-section depends closely on Z^2. Atomic interfacial ordering effects in SiGe superlattices and $CoSi_2/Si$ interfaces could be directly imaged [147].

Chemical composition profiles across III–V compound layers were successfully obtained with nanometer resolution by EDX analysis using a high-current 1 nm electron probe provided by a scanning transmission electron microscope with a field emission gun [148, 149]. However, it was realized that quantitative analysis can be severely influenced by electron channeling effects due to localization of the incident electron wave onto specific lattice sites. Then, the k factors used in the Cliff–Lorimer analysis change with orientation of the beam to the crystal, and with specimen composition and thickness. Moreover, phosphorus is lost from ternary and quaternary layers by a knock-on process when the electron beam energy of, for example, 200 keV is above the threshold energy of 110 keV for phosphorus displacement in InP [150]. Therefore, considerable care must be taken to ensure that a consistent method is adopted when performing such experiments.

3.5.2 Electronic Gallium Arsenide Device Structures

Electronic GaAs devices are based on FETs (field effect transistors), exploiting the higher electron mobility and drift velocity as compared to silicon, both of which improve device speed. In addition, use of semi-insulating GaAs substrate material minimizes parasitic capacitances. Disadvantages are the lack of a stable passivating native oxide (like thermally grown SiO_2 in silicon technology) and the higher defect density in the substrate (see Sec. 3.2.2 of this Chapter). Since both CMOS and, especially, bipolar transistors in silicon technology have improved drastically in speed in recent years, complex digital circuits on GaAs are not of general industrial interest. Analog ICs based on GaAs technology find their application at gigahertz frequencies. Microwave and millimeter wave power devices or low-noise amplifiers provide performance not obtainable with silicon. MESFETs (metal–semiconductor FETs) employ Schottky contacts as gate and ohmic contacts for source and drain metallization [151]. Process technology is in principle analogous to silicon IC technology with ion implantation for doping, lithography plus reactive ion etching for patterning, and silicon nitride layers deposited by CVD for passivation. All processes, however, have to be optimized according to the specific requirements of GaAs technology which—as with silicon technology—calls for detailed characterization during process development. Again, SEM and TEM cross-sections play a major part.

As an example, GaAs HEMTs are described which find wide application as, for example, low-noise transistors in receiver front ends for television satellite broadcasting. A HEMT is a modified form of a MESFET employing a heteroepitaxial layer structure grown by MBE [151]. At the interface of a low band gap material (GaAs) and a high band gap material ($Al_xGa_{1-x}As$) a thin inversion layer (minority charge carriers dominate) is formed, providing a very high mobility for the 2D electron gas. The cross-section through a GaAs HEMT device shows two gate fingers at A and B in Fig. 45a. Their sources are connected by an air bridge consisting of plated gold. This air bridge was deposited on a resist, which was dissolved afterwards, and

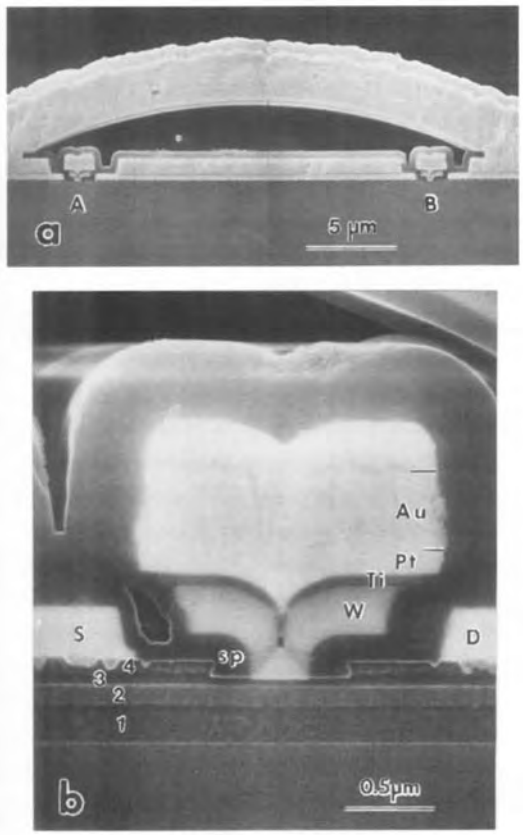


Figure 45. SEM cross-sections of GaAs HEMT device. (a) Overview with two gate fingers at A and B, and gold air bridge. (b) Gate region, for description see text.

provides a low-resistive and low-capacitance connection. The gate region is more clearly displayed in Fig. 45b. On top of the GaAs substrate, four heteroepitaxial layers can be discerned after selective etching of the SEM cross-section. Layer 1 is a $GaAs/Al_xGa_{1-x}As$ superlattice acting as a buffer layer, and the fringe pattern of the superlattice is just visible in Fig. 45b. At the interface of the undoped GaAs (layer 2) and the n-doped $Al_xGa_{1-x}As$ (layer 3) the inversion layer is formed. Using an additional strained pseudomorphic $In_xGa_{1-x}As$ layer at this interface for the channel enhances the charge carrier transport and the carrier density due to a decreased band gap, and thus improves the device properties [152]. The n^+-doped GaAs (layer 4) on top serves as the contact layer. To achieve a small gate length a self-aligned process was used for gate fabrication. First, the gate recess is etched into the contact layer using a mask of plasma silicon nitride. Then, Si_3N_4 sidewall spacers (sp in Fig. 45b) are produced by Si_3N_4 deposition and anisotropic RIE. Both RIE processes have to be optimized regarding lattice damage (see Sec. 3.3.5 of this Chapter). Using optical lithography for defining the 0.5 µm wide windows for the gate recess, the spacers narrow this window in a well-controlled way and provide a gate length of about 200 nm [153]. Tungsten is used as the gate material because of the high-temperature stability of refractory metals. To increase the conductivity of the gate, its cross-section is enhanced by additional titanium, platinum, and gold layers vapor deposited on top of the tungsten and patterned by the lift-off technique [151]. The various metal layers are indicated in Fig. 45b. All heavy elements show bright contrast in the SE image because a larger number of BSEs contribute to the SE signal. This stack of titanium, platinum, and gold is used as a mask to etch the tungsten gate. Over-etching reduces the gate width to less than that of the aforementioned stack. Then, this gate structure with a T shape ('T gate') allows self-aligned deposition of the ohmic metallization for the source and drain (S and D in Fig. 45b). The ohmic metallization further increases the thickness of the gate structure. Since ohmic metallizations involve alloying at a temperature of around 400°C (see below), a

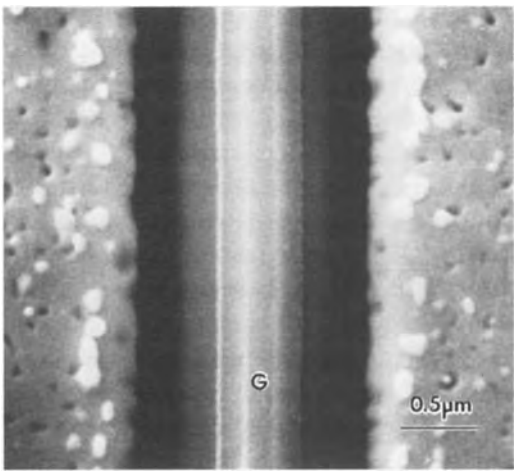


Figure 46. SEM image of the metallization-semiconductor interfaces in a GaAs HEMT after removal of the GaAs substrate by selective etching.

rough interface is formed at the source and drain contacts.

SEM cross-sections provide detailed information but only for one section plane. Since the gate fingers extend over, typically, 100 μm, assessment of uniformity along the gate requires additional information. This can be achieved by glueing the device upside down on a support and removing the semiconductor substrate and heteroepitaxial layers by selective etching. An etching solution of H_3PO_4, H_2O_2, and H_2O (13:20:5) provides high selectivity to both gate and ohmic metallization. Figure 46 shows the uniformly smooth interface of the tungsten gate (G) in the center and the rough interface of the ohmic metallization of the source and drain to the left and right.

In contrast to silicon, complex metallization systems are needed to attain ohmic contacts to GaAs. The gold–nickel–germanium system is the most widely used contact metallization for this purpose [154]. Because of its higher thermal stability, another metallization system made up of a germanium–gold–chromium–gold sequence of layers has been used successfully [155]. The complex metallurgical reactions that occur during the alloying process were studied in detail by TEM cross-sections and Auger depth profiles [156], and will be described briefly below. After vapor deposition of layers of 10 nm of germanium, 140 nm of gold, 40 nm of chromium, and 200 nm of gold, alloying was carried out at a temperature of 390°C for 12 min in a nitrogen atmosphere. The TEM cross-sections of Fig. 47 reveal the details of the alloying reaction. The morphology of the upper gold layer did not change during alloying, but the grain size has increased. Since the Auger depth profile showed the chromium layer to contain a considerable amount of oxygen, this film probably consists of chromium oxide, which also explains its very small grain size (Fig. 47b). The position of the original GaAs surface can be identified by a row of bright dots, representing oxide particles (arrows in Fig. 47b). These particles were formed when the interfacial oxide film balled up at the beginning of the alloying reaction. The grains in the vicinity of the original GaAs surface appear brighter than the gold grains, which, apart from diffraction effects, are usually darker due to their high atomic number. Individual grains were analyzed by EDX in a scanning transmission electron microscope in order to identify the various phases in the alloying zone (Fig. 47b). Only chromium and arsenic could be detected in the bright grains around the original interface with a ratio of the chromium and arsenic mole fractions close to one. The conclusion that this phase is indeed CrAs was supported by

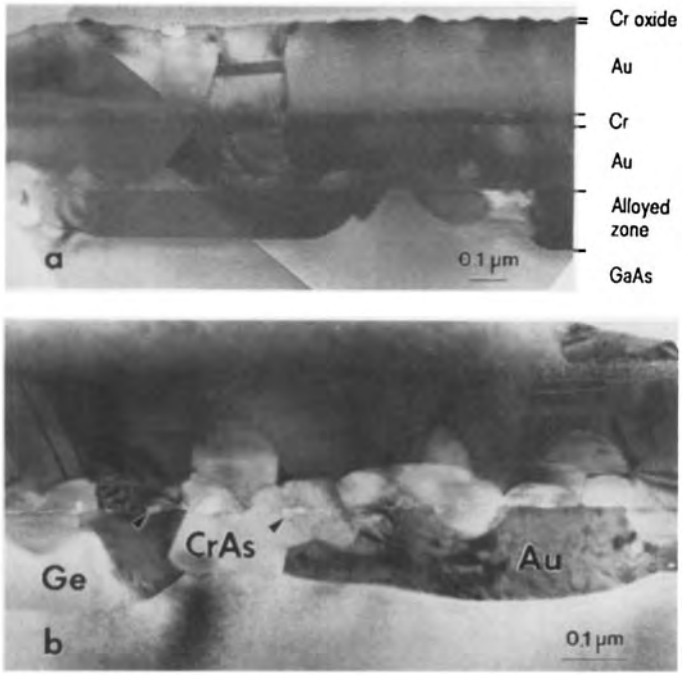


Figure 47. TEM cross-sections of ohmic metallization on GaAs after alloying. (a) Overall layer sequence. (b) The various phases in the reaction zone are indicated.

selected-area diffraction. Some grains in the alloying zone that exhibited only a slight difference in contrast from the GaAs substrate were found to consist entirely of pure germanium. These grains precipitated epitaxially onto the GaAs substrate (Fig. 47b). No significant variation was found in the composition of the gold grains in the various layers. Apart from gold, some gallium and arsenic was measured. To some extent at least, however, the X-ray intensities of arsenic and gallium stem from secondary excitation (e.g., by electrons scattered in the specimen). Since the intensity ratio of gallium to arsenic was about twice that in the GaAs substrate, we conclude that the grains of gold do in fact contain several percent of gallium. To obtain better quantitative results the upper gold layer was analyzed by electron probe microanalysis (wavelength-dispersive X-ray spectroscopy) using beam voltages low enough to confine the excitation volume to within this gold layer. No arsenic could be detected whereas the gallium mole fraction was about 7%. Furthermore, germanium was detected with a mole fraction of the order of 0.5–1%, which could not be detected by X-ray analysis in the scanning transmission electron microscope.

The line of oxide particles at the original metallization–GaAs interface (Fig. 47b) indicates that no melting occurred and alloying took place by solid state diffusion. Gold certainly provides the major driving force for the alloying reaction, supported by chromium. Gold penetrates the metal–GaAs interface and dissolves gallium, while arsenic reacts with chromium to form the CrAs phase. No compounds of germanium with other elements were found here, in contrast to the Ni_2GeAs phase described by Kuan

et al. [157]. Germanium forms a separate phase but is also dissolved in the Au(Ga) grains. Precise knowledge of the alloying reaction allows conclusions to be drawn about the conduction mechanism. During the formation of the Au(Ga) phase, gallium vacancies are produced in the adjacent GaAs crystal. The in-diffusing germanium atoms will populate gallium sites and therefore act as donors. This highly doped n^+ zone allows enhanced electron transport via tunneling. Consequently, the Au(Ga) phase that forms the major part of the contact interface is responsible for the current transport, to which neither the CrAs nor the germanium phases contribute [155, 156].

Acknowledgements

The authors express their thanks for valuable discussions and for providing literature and recent results to C. Boit, R. v. Criegen, J. Kölzer, L. Korte, R. Laubmeier, R. Lemme, G. Packesier, S. Pauthner, D. Ristow, and R. Schiessl. Further, T. Fuska is thanked for excellent photographic work, and A. Betzmeir and H. Weinel for typewriting.

3.6 References

[1] J. D. Meindl in *Proc. 3rd Int. Workshop on Measurement and Characterization of Ultra-Shallow Doping Profiles in Semiconductors* (Eds.: J. Ehrstein, R. Mathur, G. McGuire) **1995,** p. 1.1.
[2] A. W. Wieder, *Electrical Eng.* **1996,** *79*, 1.
[3] S. M. Sze, *VLSI-Technology*, McGraw-Hill, New York **1988.**
[4] S. K. Ghandhi, *VLSI Fabrication Principles*, Wiley, New York **1994.**
[5] E. Fuchs, H. Oppolzer, H. Rehme, *Particle Beam Microanalysis*, VCH, Weinheim **1990.**
[6] J. Benedict, R. Anderson, S. J. Klepeis, *Mater. Res. Soc. Symp. Proc.* **1990,** *199*, 189; ibid. **1992,** *254*, 121.
[7] A. G. Cullis et al. (Eds.), *Inst. Phys. Conf. Ser. (Proc. Microsc. Semincond. Mater.)* **1981–1995,** *60, 67, 76, 87, 100, 117, 134, 146.*
[8] F. Shimura, *Semiconductor Crystal Technology*, Academic Press, London **1989.**
[9] Z. Laczik, R. Booker, *Electrochem. Soc. Proc.* **1995,** *95–30*, 140.
[10] J. P. Fillard, P. Gall, J. Bonnafé, M. Castagné, T. Ogawa, *Semicond. Sci. Technol.* **1992,** *7*, A283.
[11] T. Abe in *Handbook of Semiconductors* (Ed.: S. Mahajan), North-Holland, Amsterdam **1994,** Vol. 3B, p. 1889.
[12] H. Föll and B. O. Kolbesen, *Appl. Phys.* **1975,** *8*, 319.
[13] H. Takeno, S. Ushio, T. Takenaka, *Mater. Res. Soc. Symp. Proc.* **1992,** *262*, 51.
[14] A. Borghesi, B. Pivac, A. Sassella, A. Stella, *J. Appl. Phys.* **1995,** *77*, 4169.
[15] H. Bender, J. Vanhellemont in *Handbook of Semiconductors* (Ed.: S. Mahajan), North-Holland, Amsterdam **1994,** Vol. 3B, p. 1637.
[16] R. W. Carpenter, I. Chen, H. L. Tsai, C. Varker, L. J. Demer, *Mater. Res. Soc. Symp. Proc.* **1983,** *14*, 195.
[17] A. Ourmazd, *Mater. Res. Soc. Symp. Proc.* **1986,** *59*, 331.
[18] R. Falster, Z. Laczik, G. R. Booker, A. R. Bhatti, P. Török, *Mater. Res. Soc. Symp. Proc.* **1992,** *262*, 945.
[19] B. Shen, T. Sekiguchi, J. Jablonski, K. Sumino, *J. Appl. Phys.* **1994,** *76*, 4540.
[20] M. Seibt, W. Schröter, *Phil. Mag.* **1989,** *A59*, 337.
[21] M. Seibt, K. Graff, *J. Appl. Phys.* **1988,** *63*, 4444.
[22] M. Seibt, *Electrochem. Soc. Proc.* **1990,** *90-7*, 663.
[23] A. S. Jordan, R. Caruso, A. R. von Neida, *Bell System Technol. J.* **1980,** *59*, 593.
[24] P. Wurzinger, H. Oppolzer, P. Pongratz, P. Skalicky, *J. Cryst. Growth* **1991,** *110*, 769.
[25] J. Weyher, J. van de Ven, *J. Cryst. Growth* **1983,** *63*, 285.
[26] A. G. Cullis, P. D. Augustus, D. J. Stirland, *J. Appl. Phys.* **1980,** *51*, 2556.
[27] O. L. Krivanek, J. H. Mazur, *Appl. Phys. Lett.* **1980,** *37*, 392.
[28] S. M. Goodnick, D. K. Ferry, C. W. Wilmsen, Z. Liliental, D. Fathy, O. L. Krivanek, *Phys. Rev. B* **1985,** *32*, 8171.
[29] Z. Liliental, O. L. Krivanek, S. M. Goodnick, C. W. Wilmsen, *Mater. Res. Soc. Symp. Proc.* **1985,** *37*, 193.
[30] A. Ourmazd, J. Berk, *Mater. Res. Soc. Symp. Proc.* **1988,** *105*, 1.
[31] H. Katsu, Y. Sumi, I. Ohdomari, *Phys. Rev. B* **1991,** *44*, 1616.

[32] M. Niwano, H. Katakura, Y. Takeda, Y. Takakuwa, N. Miyamoto, A. Hariawa, K. Yagi, *J. Vac. Sci. Technol.* **1991,** *A9*, 195.
[33] P. E. Batson, *Nature* **1993,** *366*, 727.
[34] J. M. Gibson, M. Y. Lanzerotti, V. Elser, *Appl. Phys. Lett.* **1989,** *55*, 1394.
[35] T. Ohmi, M. Miyashita, M. Itano, T. Imaoka, I. Kawanabe, *IEEE Trans. ED* **1992,** *39*, 537.
[36] H. Oppolzer, H. Cerva, C. Fruth, V. Huber, S. Schild, *Inst. Phys. Conf. Ser.* **1987,** *87*, 433.
[37] R. B. Marcus, T. T. Sheng, *J. Electrochem. Soc.* **1982,** *129*, 1278.
[38] S. M. Hu, *J. Appl. Phys.* **1991,** *70*, R53.
[39] H. Cerva, *Inst. Phys. Conf. Ser.* **1991,** *117*, 155.
[40] H. Fukuda, M. Yasuda, T. Iwabuchi, S. Ohno, *Appl. Surf. Sci.* **1992,** *60,61*, 359.
[41] R. Moazzami, *Semicond. Sci. Technol.* **1995,** *10*, 375.
[42] J. Vanhellemont, H. E. Maes, M. Schaekers, A. Armigliato, H. Cerva, A. Cullis, J. de Sande, H. Dinges, J. Hallais, V. Nayar, C. Pickering, J.-L. Stehlé, J. Van Landuyt, C. Walker, H. Werner, P. Salieri, *Appl. Surf. Sci.* **1993,** *63*, 45.
[43] H. Oppolzer, *Inst. Phys. Conf. Ser.* **1998,** *93*, 73.
[44] H. Cerva, T. Hillmer, H. Oppolzer, R. v. Criegern, *Inst. Phys. Conf. Ser.* **1987,** *87*, 445.
[45] R. M. Anderson, J. Benedict, P. Flaitz, S. Klepeis, *Inst. Phys. Conf. Ser.* **1993,** *134*, 221.
[46] T. Kamins, *Polycrystalline Silicon for Integrated Circuit Applications,* Kluwer, Boston **1988.**
[47] H. Cerva, H. Oppolzer, *Springer Proc. Phys. 35*, **1989,** p. 354.
[48] J. Washburn, *Mater. Res. Soc. Symp. Proc.* **1981,** *2*, 209.
[49] H. Oppolzer, R. Falckenberg, E. Doering, *Inst. Phys. Conf. Ser.* **1981,** *60*, 283.
[50] C. R. M. Grovenor, P. E. Batson, D. A. Smith, C. Wong, *Phil. Mag.* **1984,** *A50*, 409.
[51] H. Oppolzer, W. Eckers, H. Schaber, *J. Phys.* **1985,** *46* (Suppl. 4), C4-523.
[52] S. P. Murarka, *Metallization—Theory and Practice for VLSI and ULSI*, Butterworth-Heinemann, London, **1993.**
[53] D. Pawlik, H. Oppolzer, T. Hillmer, *J. Vac. Sci. Technol.* **1984,** *B5*, 492.
[54] R. Beyers, R. Sinclair, *J. Appl. Phys.* **1985,** *57*, 5240.
[55] R. Burmester, H. Joswig, A. Mitwalsky, *Proc. ESSDERC* (Eds.: A. Heuberger, H. Ryssel, P. Lange), Springer-Verlag, Berlin **1989,** p. 223.
[56] V. Probst, H. Schaber, H. Kabza, A. Mitwalsky, L. Van den hove, K. Maex, *J. Appl. Phys.* **1991,** *70*, 693.
[57] A. Mitwalsky, H. Körner, V. Huber, *Appl. Surf. Sci.* **1991,** *53*, 62.
[58] D. Allred, S. Jäckel, C. Mazuré, H. J. Barth, H. Cerva, W. Hösler, *J. Vac. Sci. Technol.* **1989,** *B7*, 505.
[59] W. Pamler, J. Mathuni, *J. Vac. Sci. Technol.* **1994,** *A21*, 12.
[60] A. Benninghoven, F. G. Rüdenauer, H. W. Werner, *Secondary Ion Mass Spectrometry*, Wiley, New York **1987.**
[61] H. Zeininger, R. v. Criegern in *Secondary Ion Mass Spectrometry (SIMS VII)* (Ed.: A. Benninghoven), Wiley, Chichester **1990,** p. 419.
[62] T. T. Sheng, R. B. Marcus, *J. Electrochem. Soc.* **1981,** *128*, 881.
[63] L. Gong, S. Petersen, L. Frey, H. Ryssel, *Nucl. Instrum. Methods Phys. Res.* **1995,** *B96*, 133.
[64] M. C. Roberts, K. J. Yallup, G. R. Booker, *Inst. Phys. Conf. Ser.* **1985,** *76*, 483.
[65] H. Cerva, *J. Vac. Sci. Technol.* **1992,** *B10*, 491.
[66] J. Ehrstein, R. Mathur, G. McGuire (Eds.), *Proc. 3rd Int. Workshop on Measurement and Characterization of Ultra-Shallow Doping Profiles in Semiconductors* **1995.**
[67] A. C. Diebold, M. Kump, J. J. Kopanski, D. G. Seiler in *Proc. 3rd Int. Workshop on Measurement and Characterization of Ultra-Shallow Doping Profiles in Semiconductors* (Eds.: J. Ehrstein, R. Mathur, G. McGuire) **1995,** p. 2.1.
[68] P. Muralt, *Appl. Phys. Lett.* **1986,** *49*, 1441.
[69] P. DeWolf, J. Snauwaert, T. Clarysse, W. Vandervorst, L. Hellemans, *Appl. Phys. Lett.* **1995,** *66*, 1530.
[70] A. Erickson, L. Sadwick, G. Neubauer, J. Kopanski, D. Adderton, M. Rogers, *J. Electron. Mater.* **1996,** *25*, 301.
[71] H. K. Schink, H. Rehme, *Electron. Lett.* **1983,** *19*, 383.
[72] H. J. Leamy, *J. Appl. Phys.* **1982,** *53*, R51.
[73] R. v. Criegern, F. Jahnel, M. Bianko, R. Lange-Gieseler, *J. Vac. Sci. Technol.* **1994,** *B21*, 234.
[74] R. v. Criegern, H. Cerva, G. Hobler, F. Jahnel, R. Lange-Gieseler, unpublished results.
[75] B. O. Kolbesen, H. P. Strunk in *VLSI Electronics Microstructure Science* (Ed.: G. Einspruch), Academic Press, New York **1985,** Vol. 12, p. 143.
[76] B. O. Kolbesen, H. Cerva, F. Gelsdorf, G. Zoth, W. Bergholz, *Electrochem. Soc. Proc.* **1991,** *91–9*, 371.
[77] H. Mikoshiba, N. Nishio, T. Matsumoto, H. Kikuchi, T. Kitano, H. Kaneko, *Mater. Res. Soc. Symp. Proc.* **1992,** *262*, 629.
[78] C. Claeys, J. Vanhellemont, *J. Cryst. Growth* **1993,** *126*, 41.
[79] B. O. Kolbesen, M. Dellith, R. Booker, H. Cerva, F. Gelsdorf, W. Bergholz, *Solid State Phenom.* **1993,** *32–33*, 231.

[80] W. Bergholz, G. Zoth, F. Gelsdorf, B. O. Kolbesen, *Electrochem. Soc. Proc.* **1991,** *91–9*, 21.
[81] H. Wendt, H. Cerva, in *Proc. 19th European Solid State Device Research Conf.*, Springer-Verlag, Berlin **1989,** p. 306.
[82] H. Cerva, H. Wendt, *Mater. Res. Soc. Symp. Proc.* **1989,** *138*, 533.
[83] H. Cerva, H. Wendt, *Inst. Phys. Conf. Ser.* **1989,** *100*, 587.
[84] H. Wendt, H. Cerva, V. Lehmann, W. Pamler, *J. Appl. Phys.* **1989,** *65*, 2402.
[85] J. Vanhellemont, S. Amelinckx, C. Claeys, *J. Appl. Phys.* **1987,** *61*, 2170; ibid. **1987,** *61*, 2176.
[86] J. Vanhellemont, J. Claeys, *J. Appl. Phys.* **1988,** *63*, 5703.
[87] L. Van den Hove, J. Vanhellemont, R. Wolters, W. Claassen, R. De Keersmaecker, G. Declerck in *Advanced Materials for ULSI* (Eds.: M. Scott, Y. Akasaka, R. Reif), Electrochemical Society, Pennington, NJ **1988,** p. 165.
[88] P. M. Fahey, S. R. Mader, S. R. Stiffler, R. L. Mohler, J. D. Mis, J. A. Slinkman, *IBM J. Res. Develop.* **1992,** *36*, 158.
[89] H. Wendt, S. Sauter, *J. Electrochem. Soc.* **1989,** *136*, 1568.
[90] I. DeWolf, J. Vanhellemont, A. Romano-Rodrígues, H. Norström, H. E. Maes, *J. Appl. Phys.* **1992,** *71*, 898.
[91] J. Vanhellemont, K. G. F. Janssens, S. Frabboni, R. Balboni, A. Armigliato, *Electrochem. Soc. Proc.* **1995,** *95-30*, 174.
[92] K. S. Jones, S. Prussin, E. R. Weber, *Appl. Phys.* **1988,** *45*, 1.
[93] J. P. de Souza, D. K. Sadana in *Handbook of Semiconductors* (Ed.: S. Mahajan), North-Holland, Amsterdam **1994,** Vol. 3B, p. 2033.
[94] H. Cerva, K. H. Küsters, *J. Appl. Phys.* **1990,** *66*, 4723.
[95] H. Cerva, W. Bergholz, *J. Electrochem. Soc.* **1993,** *140*, 780.
[96] M. Tamura, *Mater. Sci. Rep.* **1991,** *6*, 141.
[97] G. S. Oehrlein, *Mater. Sci. Eng.* **1989,** *B4*, 441.
[98] H. Cerva, E. G. Mohr, H. Oppolzer, *J. Vac. Sci. Technol.* **1987,** *B5*, 590.
[99] H. P. Strunk, H. Cerva, E. G. Mohr, *J. Electrochem. Soc.* **1988,** *135*, 2876.
[100] H. Cerva, H. P. Strunk in *Defect Control in Semiconductors* (Ed.: K. Sumino), Elsevier, North-Holland, Amsterdam **1990,** p. 507.
[101] M. T. Postek (Ed.), *Proc. SPIE* **1993,** *1926*.
[102] M. H. Bennet (Ed.), *Proc. SPIE* **1994,** *2196*.
[103] L. Peters, *Semicond. Int.* **1994,** March, 50.
[104] J. Frosien, E. Plies, K. Anger, *J. Vac. Sci. Technol.* **1989,** *B7*, 1874.
[105] K. M. Monahan, *Solid State Technol.* **1995,** Aug., 71.
[106] S. Nakajima, T. Takeda, *Proc. ESREF'95* **1995,** 273.
[107] *Proc. Int. Symp. for Testing and Failure Analysis (ISTFA/92,93,94)* **1992–1994.**
[108] *Proc. Eur. Symp. Reliability of Electronic Devices, Failure Physics and Analysis (ESREF 93,94,95)* **1993–1995.**
[109] E. Wolfgang, S. Görlich, J. Kölzer, *Quality Reliability Eng. Int.* **1991,** *7*, 243.
[110] *Proc. 3rd Eur. Conf. on Electron and Optical Beam Testing of Electronic Devices, Microelectronic Engineering* **1994,** *24*.
[111] J. Kölzer, J. Otto, *IEEE Design Test of Computers* **1991,** Dec., 39.
[112] J. Kölzer, E. Oesterschulze, G. Deboy, *Microelectron. Eng.* **1996,** *31*, 251.
[113] E. M. Fleuren, *IEEE/Proc. IRPS* **1983,** 148.
[114] F. Beck, *Quality Reliability Eng. Int.* **1986,** *2*, 143.
[115] V. J. Bruce, *Proc. ISTFA '94*, ASM International, Materials Park, OH **1994,** p. 73.
[116] D. L. Barton, *Proc. ISTFA '94*, ASM International, Materials Park, OH **1994,** p. 87.
[117] J. Kölzer, C. Boit, A. Dallman, G. Deboy, J. Otto, D. Weinmann, *J. Appl. Phys.* **1992,** *71*, R23.
[118] B. R. Hammond, T. R. Vogel, *Proc. Reliability Physics Symp.*, IEEE, New York **1982,** p. 221.
[119] T. Mills, *Proc. Reliability Physics Symp.*, IEEE, New York **1983,** p. 324.
[120] R. W. Belcher, G. P. Hart, W. R. Wade, *Scanning Electron Microscopy 1984/II*, SEM Inc., AMF O'Hare, IL, **1984,** p. 613.
[121] R. Lemme, R. Kutzner, S. Gonin, *Proc. ISTFA '91*, ASM International, Materials Park, OH **1992,** p. 55.
[122] H. Cerva, V. Huber, W. Eckers, A. Mitwalsky, *Ultramicroscopy* **1993,** *52*, 127.
[123] J. Orloff, *Rev. Sci. Instrum.* **1993,** *64*, 1105.
[124] K. Nikawa, *J. Vac. Sci. Technol.* **1991,** *B9*, 2566.
[125] S. Morris, S. Tatti, E. Black, N. Dickson, H. Mendez, B. Schwiesow, R. Pyle, *Proc. ISTFA '91*, ASM International, Materials Park, OH **1992,** p. 417.
[126] M. Abramo, L. Hahn, L. Moszkowicz, *Proc. ISTFA '94*, ASM International, Materials Park, OH **1994,** p. 439.
[127] P. H. Holloway, G. E. McGuire (Eds.), *Handbook on Compound Semiconductors*, Noyes, Park Ridge **1996.**
[128] M. Halliwell, *Prog. Crystal Growth Charact.* **1989,** *19*, 249.
[129] S. Mahajan (Ed.), *Handbook on Semiconductors*, North-Holland, Amsterdam **1994,** Vol. 3B.

[130] O. Ueda, *Fujitsu Sci. Technol. J.* **1992,** *28*, 1.
[131] H. Cerva, H. Oppolzer, *Prog. Crystal Growth Charact.* **1990,** *20*, 231.
[132] M. P. A. Viegers, A. F. de Jong, M. R. Leys, *Spectrochim. Acta* **1985,** *40B*, 835.
[133] E. G. Bithell, W. M. Stobbs, *Phil. Mag.* **1989,** *A60*, 39.
[134] H. Cerva, *Appl. Surf. Sci.* **1991,** *50*, 19.
[135] H. Kakabayashi, F. Nagata, *Jpn. J. Appl. Phys.* **1985,** *24*, L905; ibid. **1986,** *25*, 1644; ibid. **1987,** *26*, 770.
[136] R. W. Glaisher, D. J. H. Cockayne, *Micron* **1993,** *24*, 257.
[137] J. D. Ganiere, F. K. Reinhart, R. Spycher, B. Bourqui, A. Catana, P. Ruterana, P. A. Stadelmann, P. A. Buffat, *J. Microsc. Spectrosc. Electron.* **1989,** *14*, 407.
[138] A. Ourmazd, F. H. Baumann, M. Bode, Y. Kim, *Ultramicroscopy* **1990,** *34*, 237.
[139] S. Thoma, H. Cerva, *Ultramicroscopy* **1991,** *35*, 77.
[140] S. Thoma, H. Cerva, *Ultramicroscopy* **1991,** *38*, 265.
[141] D. Stenkamp, W. Jäger, *Ultramicroscopy* **1993,** *50*, 321.
[142] S. Thoma, H. Riechert, A. Mitwalsky, H. Oppolzer, *J. Cryst. Growth* **1992,** *123*, 287.
[143] C. Kisielowski, P. Schwander, F. H. Baumann, M. Seibt, Y. Kim, A. Ourmazd, *Ultramicroscopy* **1995,** *58*, 131.
[144] D. Stenkamp, H. P. Strunk, *Inst. Phys. Conf. Ser.* **1995,** *146*, 17.
[145] K. Ishida, *Inst. Phys. Conf. Ser.* **1995,** *146*, 1.
[146] H. Cerva, H. Riecherl, D. Bernklau, *Inst. Phys. Conf. Ser.* **1995,** *146*, 35.
[147] S. J. Pennycook, D. E. Jesson, M. F. Chisolm, *Inst. Phys. Conf. Ser.* **1991,** *117*, 27.
[148] A. J. McGibbon, J. N. Chapman, A. G. Cullis, N. G. Chew, S. J. Bass, L. L. Taylor, *J. Appl. Phys.* **1989,** *65*, 2293.
[149] N. J. Long, A. G. Norman, A. K. Petford-Long, B. R. Butler, C. G. Cureton, G. R. Booker, E. J. Thrush, *Inst. Phys. Conf. Ser.* **1991,** *117*, 69.
[150] N. J. Long, *Inst. Phys. Conf. Ser.* **1989,** *100*, 59.
[151] S. M. Sze, *High-Speed Semiconductor Devices*, Wiley, New York **1990**.
[152] T. Grave in *Devices based on Low-dimensional Semiconductor Structures. NATO Advanced Science Institute Series* (Ed.: M. Balkanski), Kluwer, Dordrecht, **1996,** p. 301.
[153] H. Schink, R. D. Schnell, *Solid-State Electron.* **1991,** *34*, 1242.
[154] N. Braslau, *J. Vac. Sci. Technol.* **1981,** *19*, 803.
[155] J. Willer, D. Ristow, W. Kellner, H. Oppolzer, *J. Electrochem. Soc.* **1988,** *135*, 179.
[156] J. Willer, H. Oppolzer, *Thin Solid Films* **1987,** *147*, 117.
[157] T. S. Kuan, P. E. Batson, T. N. Jackson, H. Rupprecht, E. Wilkie, *J. Appl. Phys.* **1983,** *54*, 6952.

Handbook of Microscopy
S. Amelinckx, D. van Dyck,J. van Landuyt , G. van Tendeloo

4 Optoelectronic Materials

4.1 Introduction

Most optoelectronic devices are based on semiconducting materials that can emit light efficiently under electrical excitation, converting electrical to optical energy (e.g. light emitting diode, semiconductor laser). Conversely, they can detect optical signals through electronic processes (e.g. photodetector) or convert light to electrical energy (e.g. photovoltaic device, solar cell).

Conventional optoelectronic devices are manufactured separately as individually packaged components, but numerous research activities have been performed recently on optoelectronic integrated circuits (OEIC) which incorporate optical and electronic devices monolithically on a single chip. This OEIC development is mainly due to a tremendous increase of semiconductor laser applications and to the progress in the very large scale integration technology (VLSI). As examples of technologies spawned by the semiconductor laser advent, there are: optical communications (long distance systems, local area networks or data transmission between satellites), laser radar and laser gyroscope, optical recording (video disks and audio disks), laser printers or displays, laser-micromachining (welding and soldering), bar code reading, medical technology (surgery and diagnostics).

Numerous industrial advantages may be attributed to OEIC including higher reliability, lower cost, higher performances and easier packaging [1, 2]. However, a difficult challenge faced by the technology is to interconnect electrically and optically various types of devices based on different materials with dissimilar structures. An illustration of this sort of problem is given by a communication system structure. The incoming electrical signal is modulated by a light source and then channeled as efficiently as possible into an optical fiber. At the end of the fiber, a detector converts the light signal back into an electrical signal. The coupling of the fiber both to the light source and to the detector is of major importance to minimize the optical losses. The transmission spectral window of the most common optical fibers presents a very low absorption at about 0.85, 1.1, and 1.5 μm and a minimum of the optical signal dispersion at 1.3 μm. The emitted wavelengths and the spectral width of the light source are thus imposed by the optical fiber properties to obtain a long distance and low cost optical communication system. In the same manner, the detectors should be designed to match their maximum sensitivity to these same wavelengths. Great

attention should be paid to the choice of the semiconductors used in such a system and to the compatibility in their mutual connection.

To date, silicon emerges as the most important material in microelectronic technology for electronic signal processing and transmission. However, due to its indirect gap nature, it cannot be intrinsically used as an efficient and universal light emitter. Concerning optoelectronic applications, the semiconductor compounds have been chosen as optoelectronic materials for their high quantum efficiency, their radiative carrier lifetime, their high surface-recombination-speed, their direct gap and the possibility of n or p-type doping to introduce efficient radiative recombination centers.

Table 1 summarizes the energy bandgaps (E_g) of various materials currently used in optoelectronic applications. III–V compounds are intrinsically good optoelectronic materials. However, their integration on silicon chips (used in microelectronic technology) is not yet fully achieved. Due to their larger bandgaps, II–VI materials offer the possibility of creating structures which operate in the visible region of the spectrum. The promise of these materials has not yet been realized because of the difficulties in producing defect-free structures with controlled doping. However, some II–VI heterojunction systems have been successfully fabricated for photodetection applications and solar cell energy conversion. Various silicon based compounds and heterojunctions have been recently studied and potential applications appear to be numerous. They are not yet developed at the industrial level because of difficulties in achieving high quality heterolayers.

It appears that most of the drawbacks listed above originate from the heteroepitaxy of active optoelectronic semiconductors on foreign substrates (e.g., silicon). Such process often causes deteriorations in epitaxial layers (high density

Table 1. Bandgap value of semiconductors most commonly used in OEIC

class	element	bandgap (eV)	direct (D) or indirect (I) gap	
IV	C	5.47	(I)	
	Si	1.12	(I)	
	Ge	0.67	(I)	
IV–VI	SiC	3.00	(I)	
III–V	GaP	2.26	(I)	
	AlAs	2.16	(I)	
	$GaAs_{1-x}P_x$	3.1 ($x = 0.45$)	(D)	$x < 0.45$
			(I)	$x > 0.45$
	GaAs	1.43	(D)	
	InP	1.35	(D)	
	InSb	0.18	(D)	
II–VI	ZnS(α) hexagonal	3.8	(D)	
	ZnS(β) cubic	3.6	(D)	
	ZnTe	2.58	(D)	
	CdS	2.53	(D)	
	CdTe	1.5	(D)	

of defects and stresses) and at the interface (misfit dislocations, etc.) due to the coupling of nonmatching lattice materials. These effects are especially detrimental for optical devices which function as minority carrier devices very sensitive to the material quality. A wide range of growth techniques has been developed in order to realize high quality heteroepitaxy including molecular beam epitaxy (MBE), chemical beam epitaxy (CBE), liquid phase epitaxy (LPE), various methods of vapor phase epitaxy (VPE) and physical vapor deposition (such as evaporation and sputtering). Concurrently performing electronic and structural techniques have also emerged in the heterolayer characterizations. Among them microscopic methods are particularly suited to the study of microstructures and their defects in very thin epitaxial layers and interfaces. This obviously covers: (i) conventional microscopy techniques such as transmission electron microscopy (TEM) and scanning electron microscopy (SEM); (ii) atomic microscopy, including scanning tunneling microscopy (STM) and atomic force microscopy (AFM); and (iii) chemical imaging techniques using ions (secondary ion mass spectroscopy), electrons (scanning Auger microscopy and electron energy loss spectrometry) and X-ray (energy dispersive X-ray spectrometry).

When specifically applied to optoelectronic materials, these microscopic techniques can provide information of the following kind:

- local composition (dopant and impurities)
- epitaxial relationships and misfit defects
- strain concentration
- morphology and microstructure.

Obviously this information should be strongly correlated to the optoelectronic properties of the materials investigated.

In this chapter, we first discuss some of the materials used as active components in the most common optoelectronic devices. The contribution of different microscopic techniques in the study of those materials is then detailed. Finally, we provide some examples of microscopic studies performed on heterojunctions or nanostructures based on III–V, II–VI and Si compounds.

4.2 Materials for Optoelectronic Devices

Materials used for optoelectronic devices can be schematically separated into two groups, depending on the physical interaction process between a radiation and the carriers of the semiconductor [3–6].

4.2.1 Luminescent Materials

Devices which convert electrical energy into optical energy are fabricated from those materials. The electrical energy applied to the semiconductor excites carriers (electrons) into the conduction band. Those electrons can fall down in free states (holes) of the valence band with '**spontaneous emission**' of a photon in the ultraviolet, visible, or infrared regions of the electromagnetic spectrum. Typical devices based on this electroluminescence phenomenon are light emitting diodes (LEDs).

The photon present in the semiconductor could also induce a transition of an

Table 2. Emission wavelength of some LED materials

material	dopant	emission wavelength (μm)	color
GaAs	Si or Zn	0.9–1.02	infrared
$GaAs_{1-x}P_x$	N	0.55–0.69	green–red
GaP	Zn or O or N	0.57–0.7	red
ZnS			blue
SiC			blue
$Ga_xIn_{1-x}As_{1-y}P_y$			infrared
$GaAs/Ga_{1-x}Al_xAs$ heterojunction			infrared

electron in the conduction band to a free state of the valence band with emission of a second photon of similar energy. This process is called '**stimulated emission**' and is used in semiconductor lasers.

The transition of carriers between the conduction and valence band is called an intrinsic band-to-band transition. The emitted light frequency corresponds then to the semiconductor gap value ($h\nu \sim E_g$). There are also other possibilities for extrinsic transitions involving carriers from one energy band to impurity levels near the opposite band (conduction band to acceptor levels, or donor levels to valence band) leading to a shorter emitted frequency ($h\nu < E_g$).

The main applications of LEDs and lasers are focused in optical coupling, optical display, and illumination.

The main requirements for a suitable LED material are: (i) a direct bandgap of appropriate width, (ii) low resistivity, (iii) efficient radiative pathway, and finally, (iv) the ease of fabrication of the material into a complete device. The most important materials used in LEDs are GaAs, GaP and ternary alloys $GaAs_{1-x}P_x$ whose characteristics are given in Table 2.

Basic principles of semiconductor lasers are quite similar to LEDs; therefore direct bandgap compound semiconductors are also required. GaAs was the first luminescent material to lase, but today most III–V compounds are extensively studied and developed in laser applications. Since the wavelength of the emitted light is very close to the width of the bandgap, wide range bandgap materials must be chosen to produce lasers which emit light in a large visible spectral width. Ternary III–V alloys of the form $A_xB_{1-x}C$ are currently used since their bandgaps vary monotonically with the value of x and lie between the AC and BC compound alloys. However, their lattice matching with the substrate is very poor. It is thus preferable to use quaternary III–V alloys which match much more closely the substrate material. Moreover, it is possible to select such alloys which are luminescent at any given wavelength from about 7 μm down to 0.55 μm. Although the two currently employed systems are $Ga_xIn_{1-x}As_{1-x}P_y/InP$ and $Al_xGa_{1-x}As_ySb_{1-y}/InP$, considerable use has also been made of the ternary IV–VI alloys PbSnTe and PbSSe, since their luminescence range varies from about 5 μm to 34 μm.

4.2.2 Photoconducting Materials

When radiation falls on the surface of a semiconducting material it can be

absorbed and carriers are generated either by a band-to-band process (intrinsic transitions) leading to the fundamental absorption or by extrinsic transitions involving impurity levels. Devices which convert optical signals into electronic processes are based on those materials. Typical examples are photodetectors and photovoltaic devices such as solar cells.

The two most common photodetectors are the photoconductive cell and the photodiode. Each type of these photodetecting devices uses specific semiconducting materials: (i) II–VI compounds under the form of polycrystalline layers are generally used for the realization of low cost photoconductive cells sensible to the visible range. CdS is conventionally used in the near infrared range but small bandgap (InSb, PbS) semiconductor compounds and Ge crystals doped with either gold or nickel are also good candidates. The $Hg_xCd_{1-x}Te$ ternary compound with a bandgap varying with x allows to realize devices with maximum sensitivity between 5 and 15 μm very useful for military applications. Table 3 compares the wavelength range for peak sensitivity of different photoconducting materials. (ii) So far, a wide range of photodiodes are fabricated with semiconductors, operating with different principles (p–i–n photodiode, avalanche photodiode, phototransistor, Schottky photodiode and heterojunction photodiode). Silicon is the major semiconducting material used in photodiode thanks to various extrinsic transitions. It provides devices with 80% quantum efficiency at wavelengths between 0.8 and 0.9 μm. Some other materials used also in photodiode technology are listed in Table 4.

Table 3. Wavelength range for peak sensitivity of different photoconducting materials

materials	wavelength (μm)
CdS, CdSe	0.3–0.7 μm
Ge (Au, Ni)	1–1.5 μm
PbS, PbSe, InSb	2 μm
$Hg_xCd_{1-x}Te$	5–15 μm (depending on x)

The simplest solar cell is based on a p–n homojunction. Incident photons pass through the thin n-type layer (called the collector) and are absorbed by the thick underlying p-type layer (called the absorber) generating carrier pairs which diffuse separately through the p–n homojunction. The solar cell performance depends on the creation efficiency of electron–hole pair and on the separation speed of these carriers across the p–n junction. Some calculated values of solar efficiency are given for various semiconductors (Table 5).

Table 4. Wavelength range for peak sensitivity of different photodiode materials

device	material	wavelength (μm)
metal–semiconductor photodiode	Ag/ZnS, Au/Si	0.3–0.7
heterojunction photodiode	$Al_xGa_{1-x}As/GaAs$	1.1–1.5
avalanche photodiode	Si	0.6–1
	Ge	1–1.6
n–i–p photodiode	Si	0.6–1
	Ge, GaAs, GaInAs, CdTe	1.1–1.5
	HgCdTe, PbSnTe	10

Table 5. Characteristics of some solar cell materials

material	bandgap (eV)	solar efficiency (%)
$CuInSe_2$	1.01	26
Si	1.12	27
Cu_2S	1.2	23
InP	1.35	29
GaAs	1.43	28
CdTe	1.56	27

4.3 Microscopic Techniques

The strong dependence of OEIC properties on the morphological and microstructural features of the used crystalline materials (and of the interfaces between layers), implies the necessity of using microcharacterization methods. These are important for checking the heterostructure quality: epitaxial relationships, crystalline defect and impurity density, interface grading, segregation of dopants, volume distribution of the chemical elements, level of strain, etc. Considering the low density level of defects (or impurities) demanded in OEIC devices, extremely high sensitivity techniques are needed. Conventional microscopic techniques (TEM and SEM) as well as atomic microscopy (AFM and STM) or chemical imaging (ionic, electronic, and X-ray mapping) are well suited for local microcharacterization of OEIC. Moreover, a correlation between atomic level information and optoelectronic device properties should be highly beneficial. For instance, cathodoluminescence (CL), electron beam induced current (EBIC) and scanning deep level transient spectroscopy (SDLTS) performed in TEM and STEM instruments are available to obtain this type of correlation.

We should point out that TEM studies necessitate difficult and tedious preparation of the samples mostly realized by mechanical thinning followed by either chemical etching or Ar^+ ion milling. The degradation possibly induced by this preparation process is one of the main drawbacks of TEM studies.

Some of the relevant questions that can be elucidated by the microscopic techniques are now discussed in detail.

4.3.1 Epitaxial relationships

At interfaces between OEIC layers, the carrier behavior is strongly dominated by interfacial states and/or electrical dipoles arising from the local chemical bonding. Therefore, the epitaxial relationships between the heterolayer and the substrate have to be well controlled. These relations, expressed as a pair of parallel directions in a pair of parallel planes, are easily determined by selected area diffraction (SAD). The comparison of diffraction patterns obtained along three orthogonal crystallographic axes gives the three-dimensional relations. One of these axes is taken parallel to the growth direction and corresponds to a plane view of the sample. The other two axes are perpendicular and correspond to transversal views.

Two cases should be distinguished depending on the constitution of the deposited layer, either large epitaxial grains with the same orientation relation or smaller grains with different orientations. In the first long-range-order case, various global diffraction techniques are available to determine the epitaxial relationships and the microscopy contribution is redundant. By contrast, the second case needs local structure determination which

can only be performed by TEM methods, by selecting very small area to obtain SAD pattern on a single grain [7]. Concurrently, local direct information is obtained by Moiré patterns (the parallel, rotational and mixed Moiré fringes can be distinguished by their relative orientation with the substrate planes) which allow the determination of the reflecting planes overlapping with a resolution of 1 nm (for a detailed description of SAD and Moiré patterns see Vol. 1 of this Handbook, Sec. 4.1.1).

4.3.2 Strain in the Heterolayer

Heterostructures are constituted by the combination of various semiconductor layers which have an artificial periodicity and structure. In such systems, the overlayers are usually limited in thickness in the range of few atomic monolayers. The epitaxial growth most often proceeds by building an overlayer strained in the growth plane to fit the substrate lattice constant. This phenomenon, known as pseudomorphism, allows buildup of strained-layer superlattices in which the thickness is small enough to accommodate the strain without creating misfit dislocations. An elastic deformation of the lattice is then observed. Quantum mechanical effects combined with strain effects drastically change the electronic band structures and can induce original optical properties of such layers [8–10].

For instance, it has been shown that the use of strained pseudomorphic heterojunctions in semiconductor lasers induces higher speed, higher modulation frequency, and lower threshold current density [11]. These characteristics are extremely sensitive to the amount of strain in the heterolayers. Two microscopic techniques are particularly well suited to determine precisely the strain content in thin layers: (i) convergent beam electron diffraction (CBED) and (ii) quantitative high resolution electron microscopy (Quanti-TEM).

Since a complete description of CBED has been presented in Vol. 1 of this Handbook, Sec. 4.1.5, we only briefly recall the advantages of this technique with regard to the strain analysis. The geometry of the High Order Laue Zone (HOLZ) lines (present inside the diffracted beams) is very sensitive to the variation of the lattice parameter and hence HOLZ patterns should be used to determine the local lattice parameter [12, 13]. Some lattice parameter variations have been measured with a precision down to $\sim 10^{-3}$. This precision in conjunction with the low diameter of the electron beam used in this technique makes it a unique means to determine local parameter changes due to elastic strains.

Quanti-TEM technique (for detailed description see Vol. 1 of this Handbook, Sec. 4.1.1) consists in measuring the variations of the projected potential over the interface area from lattice images of a heterojunction. The experimental conditions thickness/defocus are first determined using a set of calculated images. Then the intensity distribution of the experimental image is mapped. The maxima are precisely localized and related to the projection of the real structure of the sample. The measure of the exact positions of the maxima gives access to different phenomena and in particular to local elastic deformations [14–16].

4.3.3 Dopant and Impurity Volume Distribution

To modulate their extrinsic properties semiconductor materials are doped during their growth, but they may also be contaminated inadvertently with other impurities. To understand the electrical and optical properties of crystals it is necessary to control the impurity and dopant nature, their concentration, and their distribution within the layered structures. A broad range of local chemical analysis techniques have been developed and systematically applied to the new optoelectronic materials in order to gain a clearer insight into their chemical properties. Four main techniques should be cited here: microprofiling secondary ion mass spectrometry (SIMS), Auger electron spectrometry (AES), electron dispersive X-ray spectrometry (EDX) and electron energy loss spectrometry (EELS). They provide complementary information in specific areas.

For high sensitivity profiling, SIMS is the first choice method since concentration limits of $\sim 10^{13}$ atoms cm^{-3} are attainable and thus homogeneously distributed surface impurities in low concentration can be detected [17, 18]. Two types of instrumentation are used in depth profile imaging: (i) ion microprobe and (ii) ion microscope (Vol. 2 of this Handbook, Sec. 4.2.6). These techniques allow localization of chemical elements with a lateral resolution of 0.2–0.5 μm. Probing a very flat crater performed in the multilayer sample and a depth profiling image of the crater walls can be obtained and informs on the spatial distribution of the extracted secondary ions across the heterostructure thickness. For example, the electrical contacts of a transistor have been visualized by the SIMS image of Al^+ ions [19].

Higher lateral resolution in the range of 50 nm can be achieved with scanning Auger microscopy (SAM) (Sec. 2.3 of this Chapter) which is mainly limited by the incident beam diameter and the escape depth of backscattered electrons. Ultimate resolution of 10 nm has been reached with a beam energy of 5 keV and a beam intensity of 1 nA [20]. SAM is ideally suited for identifying small particulates and impurities on the wafer surface, since one can focus the electron beam on the particulate and record its Auger spectrum. Surface chemical imaging can be obtained. It is also possible to perform concentration depth profiles on a very small depth scale (nanometer range) [21].

A lateral resolution of approximately 1 nm (about the incident beam electron diameter) is obtained by EELS in TEM or STEM equipment (details in Sec. 1.3 of this Chapter). This electron spectrometry can also be used in the filtered imaging mode. This consists in integrating all the electrons which have lost an amount of energy ΔE selected by slits at the exit of the spectrometer. A procedure of background subtraction (similar to that used for SAM) allows a three-dimensional map to be obtained which informs directly on the localization of a given element. Line scan can also be performed across the interface of cross-sectional samples informing on the depth distribution of one element across the interface for example [22]. A severe limitation of the technique is in the preparation of very thin samples.

EDX can be performed in TEM or SEM apparatus. One of the advantages of using TEM is to enable a chemical map of cross-sectional thin samples to be

obtained. In this case, since beam spreading can be neglected, the spatial resolution of EDX spectrometry is approximately equal to the electron beam diameter (about 1 nm). The basic principle of the X-ray emission is detailed in Vol. 1 of this Handbook, Sec. 2.1. Briefly, the target atom ionized by the microscope electron beam produces X-ray emission associated to the inner shell level electronic transitions during the recombination process. By selecting an energy window one can obtain an elemental map of intensity. On the same sample area it is then possible to correlate the morphological characteristics of an optoelectronic structure with the spatial distribution of the chemical constituents. EDX is a quick, convenient and, at least, semiquantitative technique. However, the detection limit is very poor concerning light elements. Thus it can be used to check the presence of metallic impurities but not to detect oxides or hydrocarbon contaminants or even some kind of dopants.

4.3.4 Morphological Characterization

Morphological characterization includes the determination of layer thickness, interface and surface roughness, crystalline defect density, segregation and diffusion processes, intermediate phases at interfaces and their transitions, etc.

These morphological features are introduced during various steps of the heterolayer growth. First, for example, the clean substrate surface, which is obtained by etching of an oxidized wafer, may present structural defects like steps, vacancies, nonuniformity, and impurities capable of generating surface states. At the initial stages of deposition, the energy released by the chemisorption of adsorbed atoms can induce new reaction products and dissociate the substrate surface, generating new defects which govern the subsequent growth. Second, if now we assume that the starting substrate is perfectly flat, during the first stages of growth, surface roughening may appear because of dynamical roughness or transition from the layer-by-layer growth model to island growth mode, depending on various growth parameters.

Another important mechanism which induces grading of the interface is the segregation phenomenon. This mechanism brings substrate atoms to the topmost surface, while overlayer atoms are driven underneath, inside the substrate. This process, which takes place during the growth and sometimes at temperatures where bulk diffusion may be negligible, leads to a diffuse and intermixed interface. The interface grading can extend, even in perfectly matched layers, over several monolayers. This creates mixed chemical bonds between the interfacial atoms and their nearest neighbors in each layer and may induce significant changes in the energy levels of the heterojunction.

Furthermore, interfaces always contain interfacial defects which scatter and trap electrons, holes or excitons. These defects greatly affect the properties of optical components. In particular intrinsic defects give rise to a series of localized states well visualized in photoluminescence experiments [23]. Quanti-TEM and EELS are the most available microscopic techniques to yield information on the interface grading. EELS spectra of GaAs/AlGaAs heterostructures are given for instance in [24] and an example of Quanti-TEM

applies to Si/SiGe interface grading will be presented in the following section. The analysis of defects is commonly performed by transmission election diffraction (TED) coupled with HRTEM.

STM and AFM (detailed in Vol. 2, Sec. 7.1. and 7.2 respectively) give a direct mapping of the overlayer topography deduced from spatial variations of the electronic and atomic properties. Information on the geometry and dynamics of adsorbates can thus be easily recorded. STM (and AFM) studies mainly concern the surface reconstructions and initial stages of growth and allow the detection of local surface state defects. These surface states are of first importance for a full understanding of the growth mechanisms but they are generally not directly correlated to the optical properties of the heterostructures. However, some researchers have developed cross-sectional STM to study semiconductor heterostructures. Such studies have reported results on the local determination of band offsets, measurements of electronic transition widths at the interfaces and detailed mapping of multiple-valley conduction band structures [25].

4.3.5 Recombination Centers

We have reviewed above most microscopic techniques used for the characterization of optoelectronic materials. However, all these techniques do not allow characterization of the individual optoelectronic behavior of the chemical elements or components involved. Of particular interest is the determination of radiative/nonradiative centers in a semiconductor sample and of the wavelength of the emitted radiation. The optoelectronic properties of defects (point, line or planar) are also investigated extensively since defects play an important role in the degradation of the device optical properties.

With the advent of optoelectronic applications, new microscopic techniques have been developed. It is now possible to probe simultaneously the microstructural and the optoelectronic properties of materials with TEM or STEM microscopes equipped with cathodoluminescence (CL) systems [26–28]. Since the CL technique is introduced in Vol. 2 of this Handbook, Sec. 4.2.1 we only briefly recall the physical principle. Under the microscope primary electron beam, excitation of the target electrons from a ground state to an ionization state is followed by a relaxation process back to the ground state. During this relaxation different recombination processes take place, some of them leading to photon emission (radiative recombination). Measurements of the radiative process, which is very efficient in the case of direct bandgap materials, form the basis of CL.

Intrinsic emission results from the direct band-to-band recombination of carriers. The intrinsic emitted luminescence spectrum is often called near-band-edge radiation because of both free excitons and excitons bound to shallow recombination centers may contribute to this emission. Extrinsic emission arises from carrier transitions on shallow or deep energy states localized in the forbidden bandgap. These states associated with donor–acceptor centers result, for example, from the presence of impurities.

STEM–CL allows the measurement of the spectrum of light emitted by the sample, the display at a microscopic scale of the radiative defect centers, and the study (e.g.,

their properties) as a function of temperature. In CL spectroscopy implemented in a STEM instrument, the CL signal is collected through a very small elliptical mirror (or a lens) mounted inside the STEM polar pieces very close to the sample surface, guided by an optical fiber and detected by a photomultiplier. The signal is then analyzed by a spectrometer and sent into the video chain of the STEM microscope to obtain monochromatic CL images [20]. Some other spectroscopic techniques can be associated to CL, particularly with (i) the electron beam induced current (EBIC), which determines the recombination properties of the defects, and (ii) the scanning deep level transient spectroscopy (SDLTS), which measures the defect energy level in the gap [30].

All these spectroscopic techniques have been applied to defects and interfaces characterization in direct bandgap materials. Although unique information can be gained by a straightforward comparison of localized nanostructures observed in TEM micrographs with their radiative/nonradiative behavior, (CL) spectroscopy has not been very developed. This is mainly due to instrumental difficulties but also to the importance of the surface recombination effects. Indeed, for very thin samples there are numerous surface states with nonradiative high recombination rate that reduce the photon flux intensity.

4.4 Applications to Optoelectronic Materials

Examples of applications now follow to illustrate the contribution of some microscopic techniques for studying the different parameters, listed above, of paramount effect on the optical properties of the devices.

4.4.1 Applications to III–V Compounds

Under various growth conditions (cleaning, substrate nature, growth temperature, growth rate, etc.) III–V compound heterolayers prove to be high-efficiency luminescent materials. It is then important to clearly determine diverse origins of their light emission. These problems have been solved by microscopic cathodoluminescence techniques.

The first example [31] concerns GaAs/$Ga_{1-x}AL_xAs$ heterostructures. It compares the luminescence efficiency of two structures, a multiple quantum well (MQW) and an undoped double heterostructure (DH). The samples were grown by MBE on (1 0 0) GaAs substrates. Both photoluminescence (PL) and cathodoluminescence (CL) were measured on free-standing samples and quite similar spectra were obtained. The representative CL spectrum (Fig. 1a) obtained on a DH sample shows:

(i) a main luminescence peak at 1.513 eV which is associated with a neutral donor bound exciton (D°, X). This luminescence originates from the whole sample surface and is thus representative of an intrinsic phenomenon since nonradiative centers (at a density range of 10^4–$10^5\,cm^{-2}$) are observed in the corresponding monochromatic micrograph (Fig. 2a).

(ii) two luminescence lines associated to impurities; the carbon acceptor

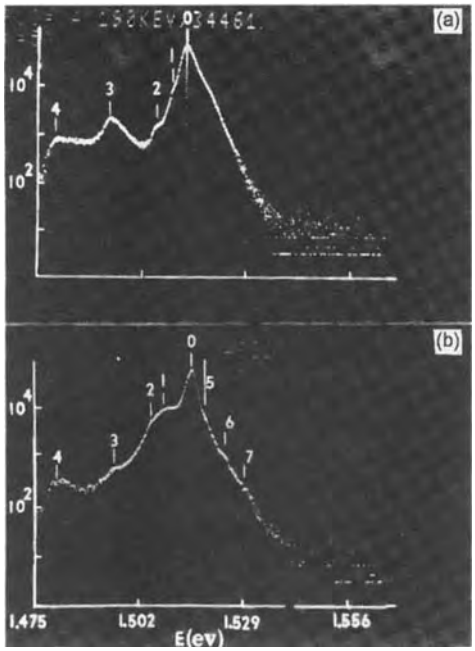


Figure 1. CL spectrum of (a) DH sample and (b) MQW sample. $E_0 = 100\,\text{keV}$. Sample temperature is 20 K. Reproduced by permission of P. M. Petroff et al. [31].

recombination (e, C°) at 1.493 eV and the germanium acceptor recombination (e, Ge°) at 1.480 eV.

(iii) two unidentified CL lines at 1.510 and 1.506 eV which seem to be associated to defects appearing during the MBE growth process. A series of CL monochromatic images in this energy range reveal that some of the discrete centers appear radiative at the energies of 1.510 and 1.506 eV (Fig. 2b), even in the highest quality areas of the sample. Concurrently, dislocation clusters or single isolated dislocation are pointed out by TEM analysis in the areas showing these discrete luminescence centers. Moreover, the disappearance of the luminescence at 4 K confirms that the defect bound exciton or

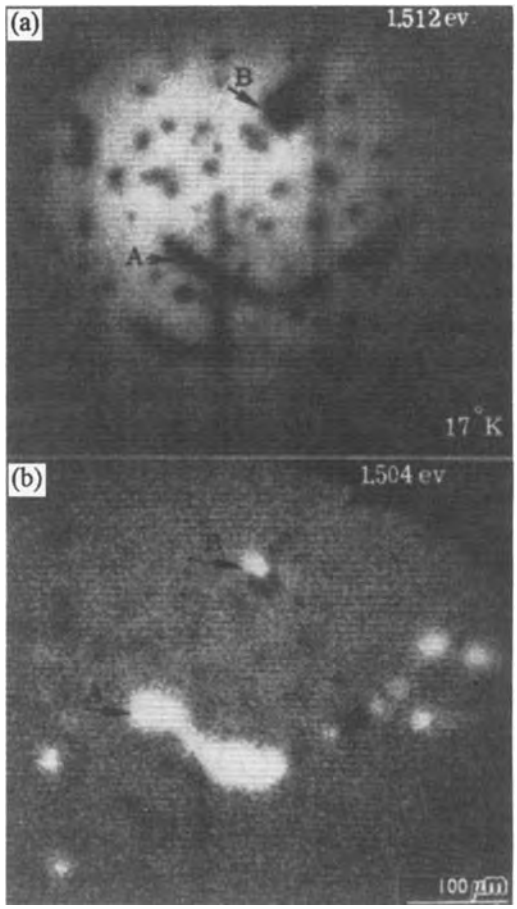


Figure 2. CL monochromatic micrograph of the DH sample at (a) 1.512 eV and at (b) 1.504 eV. Sample temperature is 17 K. Reproduced by permission of P. M. Petroff et al. [31].

electron is at the origin of this luminescence. Therefore it can be easily concluded that the 1.510 and 1.506 eV CL lines are related to the observed dislocations.

By comparison, the representative CL spectrum of a MQW structure (Fig. 1b) shows:

(i) a main luminescent line (at 1.516 eV) corresponding to the recombination energy of the free exciton attributed

to the electron–heavy hole (e–hh) $n = 1$ transition.

(ii) three emission lines at 1.519, 1.524 and 1.530 eV, respectively, attributed to the electron–light hole (e–lh) $n = 1$ transition, to the forbidden transition and to the (e–hh) $n = 2$ transition.

(iii) the below band gap luminescence which is identical to that of the DH structure. Even if the origin of the 1.513 and 1.508 eV emission lines is interpreted in the same manner than for the DH structure, it should be noticed that their intensity is much lower. Therefore, since a dislocation density of 10^4–$10^5\,cm^{-2}$ has been revealed by TEM, these are mainly inactive dislocations. Furthermore, by contrast with the DH structure, the monochromatic micrographs reveal a perfect uniformity of luminescence and the absence of nonradiative discrete centers. This result explains the higher luminescence efficiency observed in the MQW superlattices over that of the DH structures.

The second example of STEM–CL technique [32] analyzes degradations induced by the etching process on free-standing quantum dots fabricated from GaAs/AlGaAs quantum well heterolayers. The structure studied contained four single quantum wells with nominal GaAs well-widths of 5, 10, 20 and 80 nm separated by AlGaAs barrier thicknesses of 34 nm. The quantum dots were produced by a combination of electron beam lithography and reactive ion etching using CH_4/H_2 ambient. The etching depth was large enough to ensure that the 5 and 10 nm wells formed into dots. An array of dots 500 nm in diameter and 2 μm spacing was produced.

A representative spectrum recorded on unetched areas of the sample (Fig. 3a) shows peaks at 1.591, 1.543 and 1.521 eV that correspond to the quantum well excitonic emission from the 5, 10 and 20 nm

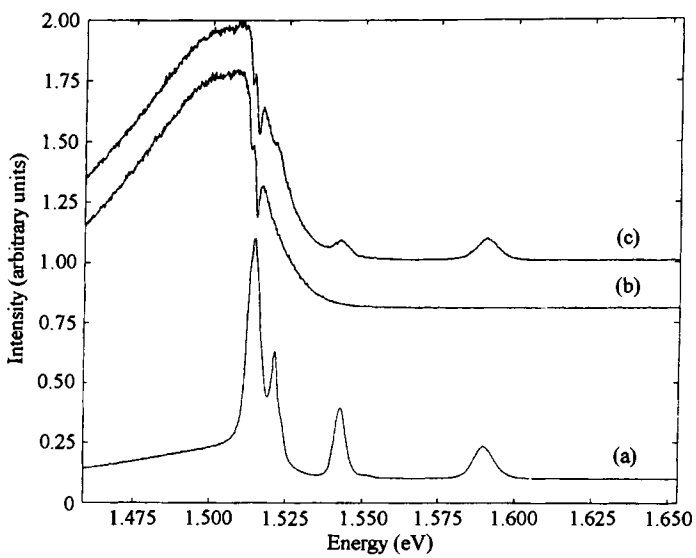


Figure 3. CL spectra from (a) unetched area, (b) etched area, and (c) individual quantum-dot. Reproduced by permission of J. Wang et al. [32].

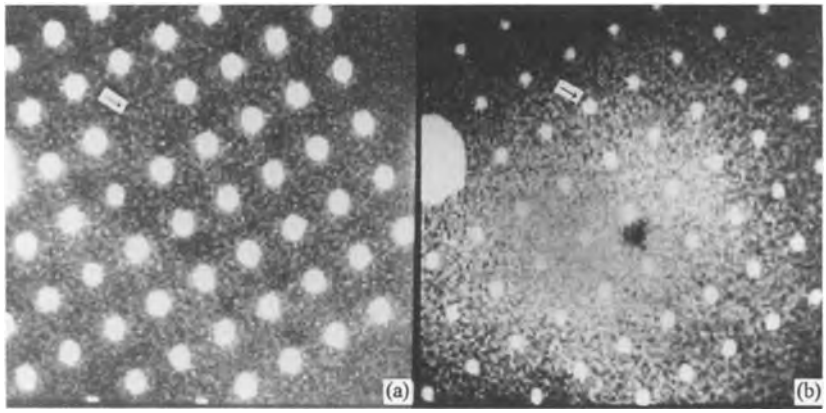


Figure 4. CL monochromatic image of the quantum dots at the emission energy of the quantum dots formed from (a) 5 nm (1.591 eV) and (b) 10 nm (1.543 eV) quantum wells. Reproduced by permission of J. Wang et al. [32].

wells respectively. A broad peak appears at 1.514 eV, which results from the emission of both the bulk substrate and the 80 nm well. The low energy tail of this peak is relative to impurities in the substrate. In spectra recorded on etched areas, only this low energy tail of the 1.514 eV peak dominates the curve (Fig. 3b); the emission lines from 5 and 10 nm wells disappear and those from the 20 and 80 nm are greatly reduced in intensity. This result shows that a high degradation of the quantum well side walls is induced by the etching process, leading to a high non-radiative surface recombination velocity.

By contrast, the representative spectrum of an individual 500 nm diameter quantum dot (Fig. 3c) shows the individual signal corresponding to the 5 and 10 nm quantum wells. The change of the effective energy gap caused by a lateral carrier confinement is unobservable in the spectrum. This result confirms the calculations which show that for quantum dots larger than 120 nm the energy shift should be lower than 1 meV. Two CL micrographs were taken at the energy corresponding to the 5 nm (Fig. 4a) and to the 10 nm (Fig. 4b) quantum wells, respectively. The dot-image/background intensity ratio is much higher for the dots formed by the 5 nm well than that for the 10 nm well. This difference can be explained by a lower carrier diffusion coefficient for the narrower quantum wells, resulting in a weaker sidewall-surface non-radiative effect. A higher degradation of a 5 nm depth dot (missing dot indicated by an arrow) should be noticed as compared to its corresponding 10 nm depth dot.

These two studies reveal the advantages of the microscopic CL technique to correlate structural-optical properties of nanostructures.

4.4.2 Applications to Si Based Compounds

About 95% of the microelectronic circuits rely on silicon chips. However, because its relatively small indirect bandgap, the band-to-band luminescence of bulk

crystalline silicon occurs in the near infrared wavelength with a very small efficiency. This is the reason for the quest of other optoelectronic semiconducting materials with higher efficiency and preferably compatible with the VLSI Si technology. The most common optoelectronic devices such as electroluminescent diodes for instance are fabricated from other semiconductors (GaAs and related compounds) but their integration in Si microcircuits still remains imperfectly resolved. Alternatively, luminescent Si based materials are also recently investigated in order to benefit all advantages of the digital Si advanced technology. Few studies of silicon based structures, with the use of microscopic techniques will be described here. The examples concern (i) luminescent porous silicon, (ii) Si/SiGe, and (iii) $FeSi_2$/Si heterostructures.

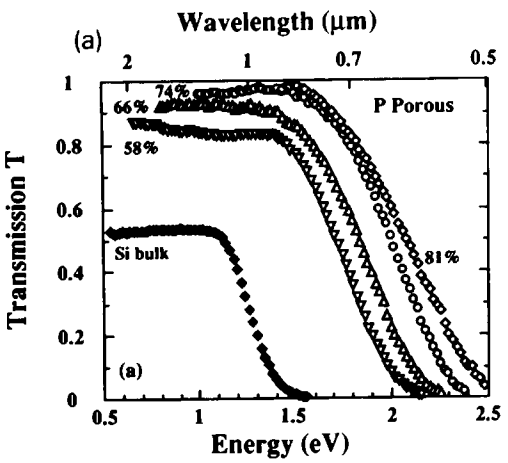
4.4.2.1 Luminescent Porous Silicon

The recent discovery that porous silicon (PS) layers can emit efficiently visible light at room temperature has boosted numerous investigations due to its potential applications. The generic name of PS includes various and complex materials whose properties (structural, optical, electrical, etc.) strongly depend on many different parameters. Generally PS is obtained by anodic dissolution of silicon in hydrofluoric acid solutions in an electrochemical cell with the silicon substrate biased as the cell anode [33]. The process results in a sponge-like sample constituted with an Si nanostructure skeleton (wires, dots, etc.) surrounded by empty pores.

The PS layers analyzed here were formed on (1 0 0) silicon wafers, either light (p^-) or highly (p^+) doped p-type. It has been shown that when the porosity of PS exceeds about 50% the samples display visible luminescence. Although the exact origin of this phenomenon is still controversial, the proposed mechanisms prevalent in the literature are centered around quantum size effects and surface passivation. This study compares both the photoluminescence and the microstructure of high porosity p^- and p^+ luminescent PS (LPS).

Optical transmission values of different porosity LPS samples (p^- or p^+ doped) are presented in Fig. 5. The increase in the bandgap energy E_g for the 80% porosity samples (p^- and p^+) can be explained by quantum confinement effect in the small silicon crystallites which constitute the silicon skeleton. The calculated values of the crystallites dimensions (L) are given simply by $E_g = h^2/4m^*L^2$ where h is the Planck's constant, m^* the effective electron or hole mass. L is in the range 2–4 nm in the case of lightly doped samples and 7–9 nm in the case of highly doped samples. Photoluminescence (PL) spectra of p^+ and p^- LPS are shown in Fig. 6. Results of a detailed comparative study reveal that spectra of both samples present a maximum at about the same wavelength. PL intensity of p^+ samples is however much less intense. When the porosity of the layer increases (L decreases) one observes a blue shift of the PL peak, which is in agreement with the quantum confinement estimation.

Concurrently, TEM analysis has been performed on those LPS samples ([34] and references therein). Although the preparation process for TEM analysis partly degrades the samples, valuable morphological and microstructural information can be gained. A detailed description of



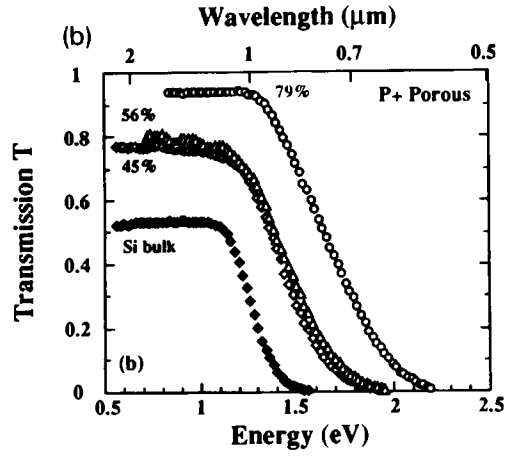


Figure 5. Transmission coefficient versus photon energy for 40 μm thick free-standing PS and substrate doping (a) p^- and (b) p^+ samples. Reproduced by permission of G. Bomchil et al. [33].

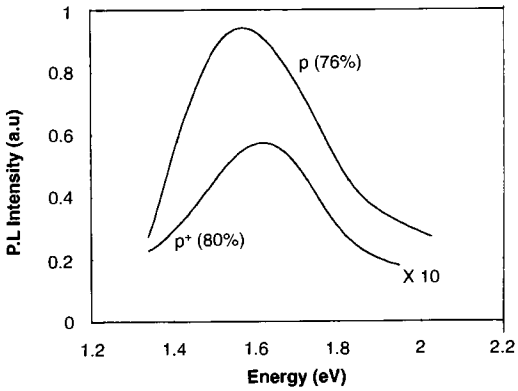


Figure 6. Photoluminescence intensity versus wavelength for p^- and p^+ samples. Reproduced by permission of G. Bomchil et al. [33].

information provided by different TEM associated techniques can be found in [35]. When the porosity is increased from 60 to 65% a decrease of the crystallite size from the 10–15 nm range to the 5–8 nm range is observed on the TEM micrographs. Moreover, HRTEM images show that the 80% (p^-) PS is formed with an array of fine holes and crystallites, while TED exhibits large spot streaking and diffuse scattering (Fig. 7). Two phases are clearly seen: (i) crystalline clusters with dimensions ranging from 1.5 to 4 nm. Various degradations (undulations and distortions) of the crystalline lattice of these clusters are observed (Fig. 8); (ii) an amorphous like contrast giving rise to broad diffuse rings. Clusters misoriented from each other with less than three superimposed planes or amorphous silicon could explain this contrast.

By contrast, 85% (p^+) LPS layers exhibit very different morphology. They rather display an anisotropic structure of a longitudinal-branched network of pores/silicon rods (Fig. 9). From HRTEM analysis two crystalline components should be distinguished (Fig. 10): (i) long monocrystalline rods with 10 nm or more diameter, and (ii) small rotated silicon crystallites with size ranging from 2 to 6 nm.

A quantitative estimation of the clusters density is rather difficult. However, a direct comparison between p^- and p^+ PS samples proves that many fewer crystallites with size in the range 2–4 nm are

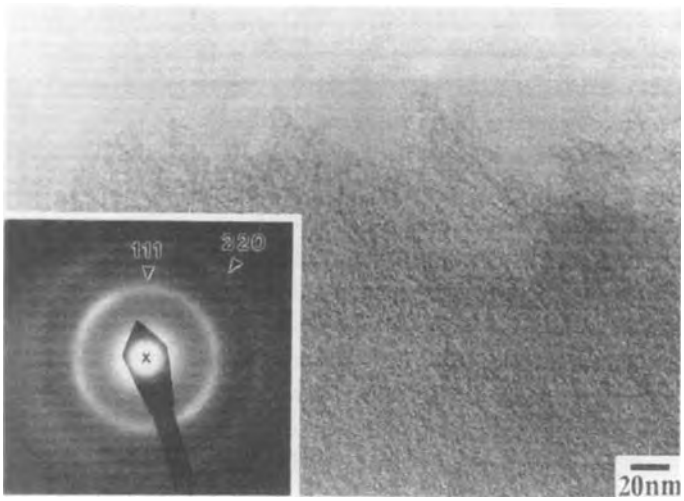


Figure 7. Overall view of 85% LPS p^- doped and corresponding TED pattern. Selected area diaphragm diameter: ~10 μm. Reproduced by permission of I. Berbezier et al. [34].

present in the latter case leading to a less intense PL signal. This result shows the influence of crystallites of very small size on the luminescence properties of LPS.

Since the various LPS microstructures generally observed should lead to different optical properties, the origins of the luminescence could be elucidated by systematic comparative TEM/PL studies.

Characterization at the atomic level could also be performed by EELS (together with HRTEM) which offers higher spatial resolution chemical analysis using a STEM. These techniques have been applied to investigate the chemical bonding of n^+-type LPS [36]. Samples were formed by electrochemical oxidation under anodic polarization of silicon in aqueous HF electrolyte assisted by illumination (photoelectrochemical etching). As shown by HRTEM, n^+-type LPS presents quite different morphology than a p-type

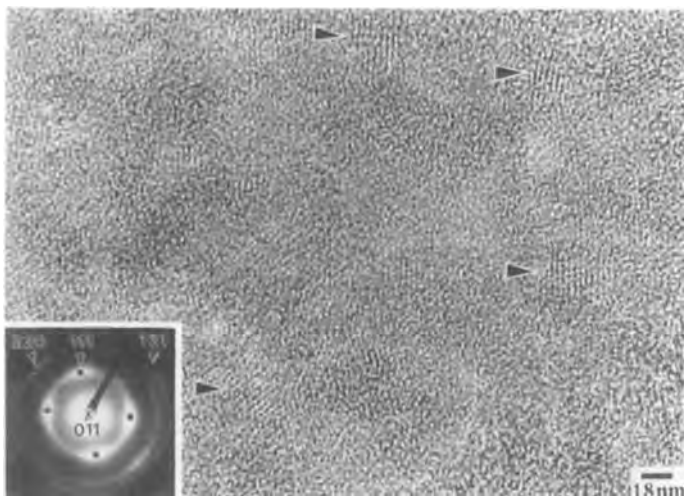


Figure 8. HRTEM and corresponding TED pattern of 85% LPS p^- doped sample. Selected area diaphragm diameter: 0.5 μm. Reproduced by permission of I. Berbezier et al. [34].

Figure 9. Overall view of 85% LPS p^+ doped sample. Reproduced by permission of I. Berbezier et al. [34].

one. n^+-type LPS consists in wires composed of a crystalline core surrounded by a layer exhibiting typical granular contrast of noncrystalline material. The thickness of this surrounding layer was estimated to 1.5–2 nm. The core of the tangled silicon wire exhibited various diameter sizes ranging from 3 to 20 nm which often undulated in width. The determination of the nature of the amorphous coverage has been performed by EELS. Two energy loss regions of interest were scanned to obtain information on the electronic structure of the silicon wire: the low loss region and the fine structure occurring over the 10–30 eV above the Si L_{23} ionization edge.

Data acquisition consisted in a set of spectra recorded at given positions separated by a distance of about 0.5 nm across the silicon wire. Data for the L_{23} ionization edge region are displayed as three dimensional intensity plot along two axes (Fig. 11a), one corresponding to the position on the sample and the other to the energy loss scale. This perspective of the spectra reveals clearly the similar shape of all the L_{23} ionization edge whatever was the probe position. A selection of different spectra is extracted from the above sequence: (i) close to the surrounding amorphous layer (Fig. 11b) and (ii) within the silicon wire, between the crystalline core and the amorphous layer (Fig. 11c). It can be seen that the shape of the L_{23} ionization edge remains constant, and that no detectable chemical shift can be observed across the whole wire.

Similar sets of data acquisition performed in the low loss region show: (i) that the energy position of the volume plasmon does not shift noticeably, as long as it can be identified, (ii) that the

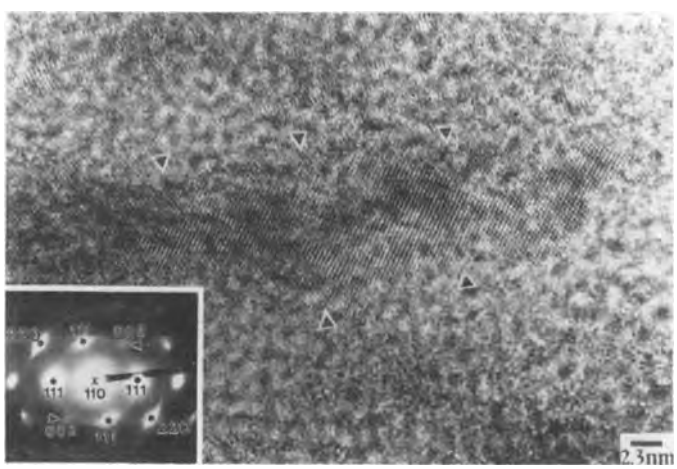


Figure 10. HRTEM and TED pattern of 85% LPS p^+ doped sample. Reproduced by permission of I. Berbezier et al. [34].

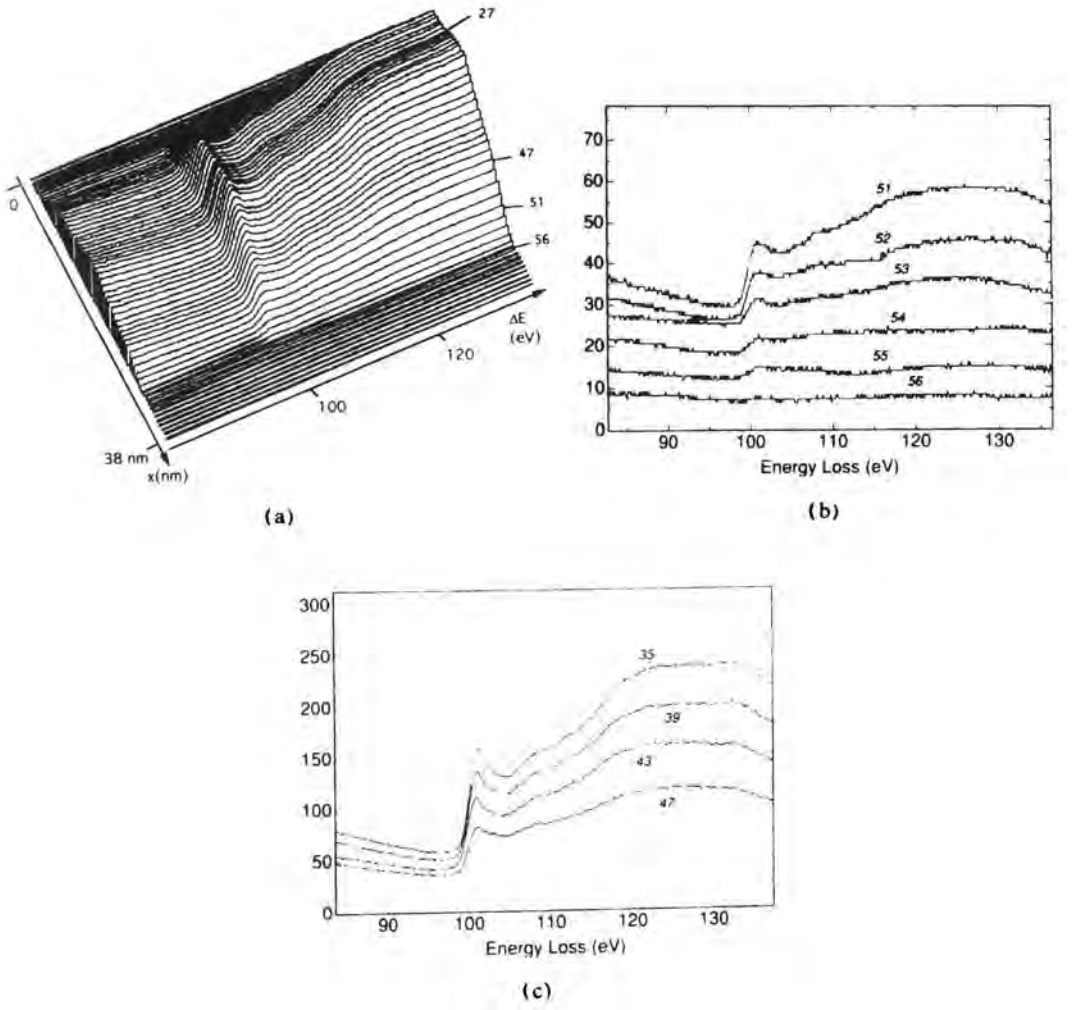


Figure 11. (a) Sequence of EELS spectra acquired as the electron probe is scanned across a single wire. It displays the behavior of the Si L_{23} edge. Selection of spectra extracted for the above sequence (b) close to the amorphous surrounding layer and (c) within the wire. Reproduced by permission of A. Alby-Yaron et al. [36].

two first structures at the onset of interband transitions are also quite identical across the whole wire.

In conclusion, EELS results [36] demonstrate the absence of foreign species (SiO_2 and SiH) in the surrounding amorphous layer and thus support the quantum confinement model to explain the luminescence of n^+-type porous silicon.

4.4.2.2 $Si_{1-x}Ge_x$/Si Strained Superlattice Structure

Similar to the III–V or II–VI related compounds, $Si_{1-x}Ge_x$ alloys display at least two attractive features. By controlling the chemical composition x [37] one can play: (i) with the lattice mismatch δ, as compared to the Si substrate ($x \in (0, 1)$,

$\delta \in (0, \sim 4\%)$) and then control the strain in pseudomorphic heterolayers and their critical thicknesses and hence the optoelectronic properties; and (ii) with the fundamental energy gap E_g ($x \in (0, 1)$, $E_g \in (1.12\text{–}0.67\,\text{eV})$).

Other advantages include the Si technology compatibility, quantum effects with intersubband transitions allowing far infra red applications, non linear optical properties. For example it would be very attractive to use $Si_{1-x}Ge_x/Si$ strained layer superlattices (SLS) for light emitting diodes or photodetectors. However, one should notice that the $Si_{1-x}Ge_x$ alloy remains an indirect gap semiconductor at all values of x and a low absorption coefficient is predicted [38]. The comparison of photoluminescence efficiency and spectra recorded on strained and relaxed layers is shown for instance in [39]. Some basic ideas on the calculations of the strain and on the critical thickness are given in [40]. Many microscopic studies have been applied to the $Si_{1-x}Ge_x/Si$ system to determine its epitaxial relationships, its dopant distribution in volume and at interfaces, its strain and dislocation density, its morphology, etc. Here we will only detail a determination of the interface roughness by Quanti-TEM. This morphological detail is of great interest in the heterostructure buildup since it governs most of the optoelectronic properties of the material. This image analysis can also be applied to the strain determination.

In Quanti-TEM, the variation of the potential in crystalline solids is measured in any projection, without knowledge of the imaging conditions (detailed description in [41]). Image analysis [42] is then applied to determine the chemical composition mapping from an experimental image. The latter is divided into unit cells, corresponding to the lattice cells, the intensity distribution of which is digitized and represented by a vector R^i. The experimental image potential is then related to the projected crystalline potential of the material under specific imaging conditions (defocus, thickness of the sample). The R^i vectors are determined over a region of the sample with a chemical composition of reference. Changes in composition induce changes in the path of R^i and also in the rate at which the path is traversed in each material. Figure 12 shows the HRTEM image of a $Si_{0.75}Ge_{0.25}/Si$ interface and Fig. 13 is the composition map across the quantum well obtained by Quanti-TEM after thickness normalization. This quantitative microscopic map of the compositional change reveals the large extent of intermixing over the interface not visible on the HRTEM image. All these results should be taken

Figure 12. Lattice image of $Si/SI_{0.75}Ge_{0.25}/Si$ quantum well structure viewed in $\langle 1\,1\,0 \rangle$ cross-section. Reproduced by permission of P. Schwander et al. [42].

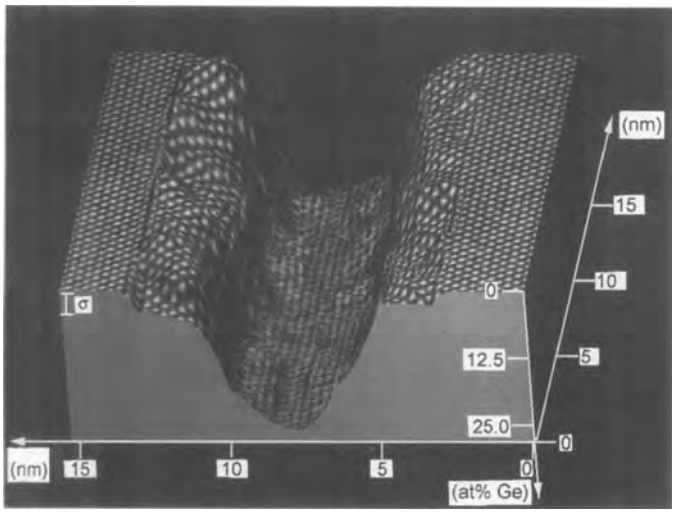


Figure 13. Quanti-TEM composition map of $Si/Si_{0.75}Ge_{0.25}/Si$ quantum well structure. Heights represent Ge concentration. Reproduced by permission of P. Schwander et al. [42].

into account during the optical properties calculation.

4.4.2.3 $FeSi_2$/Si Interface

Recently the semiconducting silicide phase β-$FeSi_2$, which displays an optical gap close to 0.87 eV, has been extensively studied since it may be a very attractive material for novel integrated optoelectronic devices [43].

Due to the crystallographic properties it is possible to epitaxially grow orthorhombic β-$FeSi_2$ phase on silicon (0 0 1) and (1 1 1) surfaces and indeed numerous experimental studies achieved thin β-$FeSi_2$ epitaxial layers by different deposition modes [43].

Beside the stable β-$FeSi_2$ phase grown so far with solid phase epitaxy (SPE) and reactive deposition epitaxy (RDE), surprising results have been obtained. In particular, the appearance of various and unexpected $FeSi_2$ phases preceding the formation of the stable β-$FeSi_2$ has been observed. Among these 'new' $FeSi_2$ phases, two cubic metastable phases (not present in the thermodynamic phase diagram) strained and stabilized by epitaxy on silicon substrate have been identified by different groups of researchers (hereafter call CsCl- and γ-$FeSi_2$ phase). Moreover, during the RDE growth experiments at low temperature (~550°C), the metallic tetragonal α-$FeSi_2$ phase has been revealed. This phase thermodynamically stable only at high temperature (above 940°C) has been surprisingly stabilized during the deposition process at low temperature [44]. The appearance of all these metastable phases, not thermodynamically foreseen, leads to a rather confusing situation in the literature. Their identification is a rather difficult problem for at least two reasons: (i) they are observed only at very low thicknesses (a few nanometers), and (ii) when they grow on Si substrates their crystallographic structures are very close and give rise to similar diffraction pattern. Therefore, they call for a thorough knowledge of each phase in order to make sure their distinction.

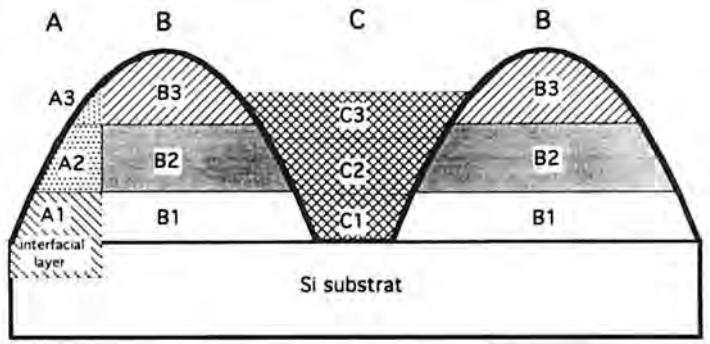


Figure 14. Schematization of the areas scanned and analyzed by image analysis. Reproduced by permission of I. Berbezier et al. [45].

Taking into account these remarks we have presented [45] a complete transmission electron microscopy (TEM) and transmission electron diffraction (TED) characterization of thin α-$FeSi_2$ films obtained by RDE deposition mode, on top of Si(1 1 1) surface. As an illustration of a complex multiphase-interface we reproduce, in Fig. 14, a HRTEM image of an iron disilicide/silicon interface.

The interface is atomically plane on a 20 nm mean averaged length. Macrosteps with spacing to each other of about 20–30 nm are observed with interfacial dislocations confined at the edge of these steps in agreement with the calculated distance between misfit dislocations (only due to the mismatch between the (2 2 0) Si and the (1 1 0) α-$FeSi_2$).

More quantitative information can be provided by performing a careful analysis of HR images. The (2 2 0) silicon planes separated 0.192 nm, are used to calibrate the distances on each image. The interplanar distances are averaged over a hundred planes leading to a measurement accuracy of ± 0.001 nm.

About twenty islands have been studied and, for each island, different areas are systematically scanned along a line perpendicular to the interface. Three regions have been probed: (i) in the near vicinity of the interface (within 1 nm from the interface), (ii) at approximately 2 nm from the interface, and (iii) on the top of the $\sim$500°C grown α-$FeSi_2$ film.

Parallel to the silicon surface three typical interfacial growths (A, B, and C) are observed (Fig. 15). The (1 1 0) α-$FeSi_2$ interplanar distances measured for each growth (A, B, and C) along the line normal to the silicon substrate surface are noted A_1, A_2, A_3, B_1, B_2, B_3 etc. starting from the

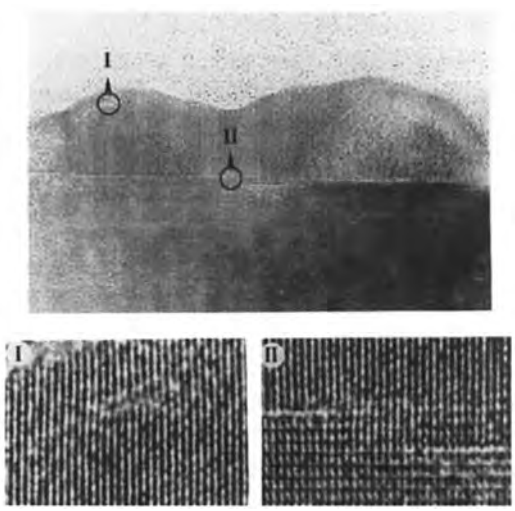


Figure 15. Cross-sectional analysis of an $FeSi_2$ grain along the ⟨1 1 2⟩ Si axis. Dislocations have been observed close to the silicide surface (area I) and at the interface (area II). Reproduced by permission of I. Berbezier et al. [45].

Table 6. Interplanar distances measured in different areas of the epitaxial film

area	measured (nm)	crystallographic data
A_1	0.192	(2 2 0) Si
A_2	0.189	(1 1 0) relaxed α-$FeSi_2$
A_3	~0.196	(0 0 1) or (0 1 0) β-$FeSi_2$
B_1	0.192	(2 2 0) Si
B_2	0.189	(1 1 0) relaxed α-$FeSi_2$
B_3	~0.196	(0 0 1) or (0 1 0) β-$FeSi_2$
C_1–C_3	0.189–0.196	From relaxed α and β-$FeSi_2$ phase

interface towards the topmost surface of the α-$FeSi_2$ layer. These are reported in Table 6 and compared to the crystallographic data of the relevant silicide and silicon interplanar distances in order to identify them.

Our HR measurements may be summarized as follows. In areas A (Fig. 15) at the Si substrate interface one first observes a very thin cubic $FeSi_2$ layer totally coherent with the Si substrate ($d \approx 0.192$ nm measured at A_1). It evolves to a relaxed tetragonal α-$FeSi_2$ phase ($d \approx 0.188$ nm measured at A_2) which, at its turn, transforms to the stable orthorhombic β-$FeSi_2$ phase ($d \approx 0.196$ nm measured at A_3). One remarks that the cubic/α-$FeSi_2$ interface is atomically abrupt while the relaxed α/β-$FeSi_2$ interface is usually broad, containing dislocations and disordered areas, reflecting faithfully the similarity of the interfacial structures between cubic and α-phases and in contrast the difference between the α and β-$FeSi_2$ ones.

An alternative of the $FeSi_2$ epitaxial growth is also observed (areas B, Fig. 15). In this case another strained α-$FeSi_2$ layer coherently covers the Si substrate ($d \approx 0.192$ nm measured at B_1). It replaces and plays the role of the cubic strained layer of the first case. This strained α-layer evolves towards a relaxed α-phase ($d \approx 0.188$ nm measured at B_2) and then towards the β-$FeSi_2$ phase ($d \approx 0.196$ nm at B_3). In those B areas at the Si interface, macrosteps are observed with interfacial dislocations, even though the α-layer at intimate contact with the Si substrate is not relaxed but still coherently strained (areas B_1). It only relaxes a little bit further from the interface (areas B_2). We believe that our observations are reflecting the initial stages of a strained–relaxed layer growth mediated through an interfacial dislocation formation.

Some areas called C in Fig. 15 display a quasicontinuous variation of lattice parameter which increases gradually from 0.188 nm (measured at C_1) corresponding to the relaxed α-phase, and towards 0.193 nm (measured at C_3) close to β-$FeSi_2$.

Our results shed light on the $FeSi_2$ epitaxy mechanisms. At the very intimate contact with the Si substrate coherent ultrathin $FeSi_2$ films should grow. This leads to either a pseudomorphic cubic phase or to a strained α-$FeSi_2$ phase. The following transformation of these phases to the stable orthorhombic β-$FeSi_2$ passes through a fully relaxed tetragonal α-$FeSi_2$ phase. Obviously, the growth is strongly dependent on uncontrolled factors since at

least two kinetic paths have been shown. The above mentioned uncontrolled factors may include mainly kinetic factors (depending on the growth technique, the growth temperature and the initial substrate morphology). The reasons why cubic or/and strained tetragonal $FeSi_2$ phase can easily grow on top of silicon might be attributed to the influence of the strains exerted by the silicon substrate during the epitaxial growth.

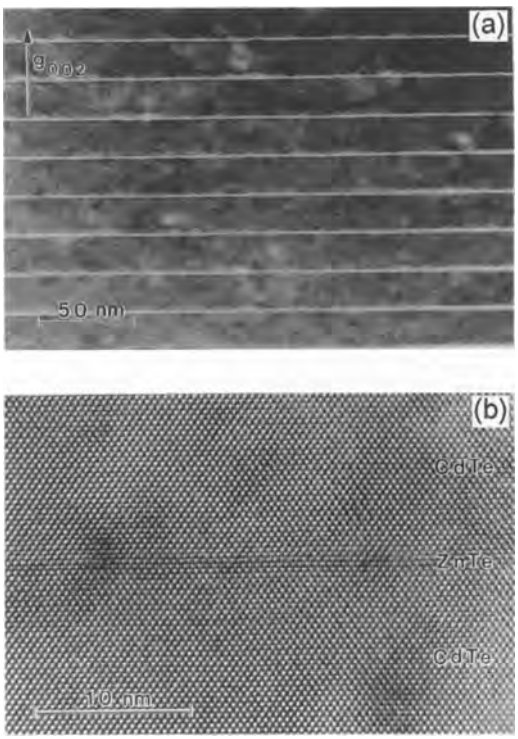


Figure 16. (a) Dark field imaging and (b) HRTEM of 57/2 CdTe/ZnTe samples. Reproduced by permission of P. H. Jouneau et al. [46].

4.4.3 Application to II–VI Compounds

$CdTe/Cd_{1-x}Zn_xTe$ is one of the most interesting II–VI heterostructures because of the valence band discontinuity induced by strain. Strained multiquantum wells (MQW), with strain orientated along polar crystalline direction present a high piezoelectric field. This field modifies the band structure and thus the radiative recombination energy. Since the strain distribution determines the localization and the intensity of the piezoelectric field, its characterization with TEM is of utmost importance as demonstrated below [46].

The $CdTe/Cd_{1-x}Zn_xTe$ MQW were grown by MBE with experimental conditions adjusted to avoid the strain relaxation of the well layers and to keep the average strain of the MQW close to zero (strain symmetrization). TEM analysis reveals the absence of misfit dislocations attesting the pseudomorphic growth and allows to determine the MQW period. Illustration of the 57CdTe/2ZnTe structure is given in Fig. 16 for which (a) displays a bright-field micrograph and (b) the corresponding HRTEM image. By digitizing HRTEM and with the use of an image processing software, local distortions can be measured. Figure 17 shows values of local distortion measured on three different samples (27/1, 57/2, and 88/3 periods, respectively) and by comparison the distortion calculated from the continuum elastic theory. It can be seen that the calculated values do not fit exactly the experimental curves. Other adapted models are proposed by the authors. About 70% of the strain is accommodated in the ZnTe layer and that residual strain extends over 3–5 ML after the last ZnTe plane location. Zn segregation, revealed by the asymetric strain profile, induces this residual strain.

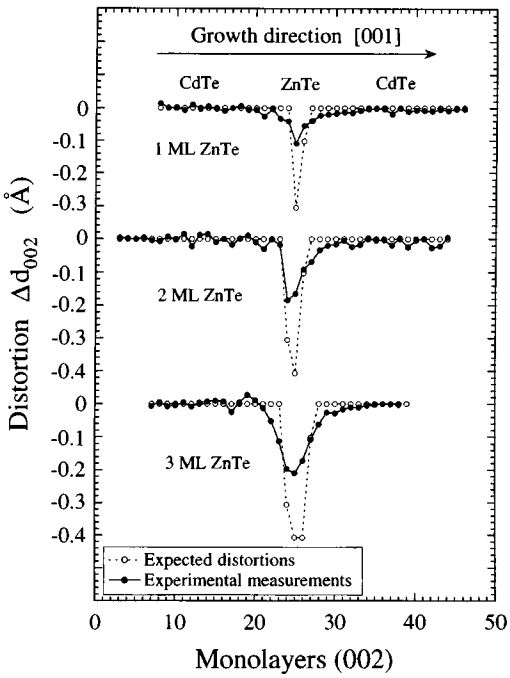


Figure 17. Distorsion analysis results deduced from HRTEM of 27/1, 57/2 and 88/3 CdTe/ZnTe samples. Reproduced by permission of P. H. Jouneau et al. [46].

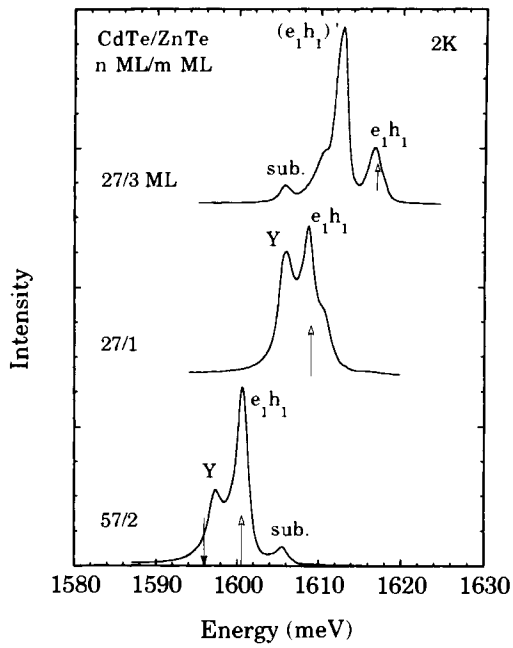


Figure 18. PL spectra of 27/1, 27/3 and 57/2 CdTe/ZnTe samples. Reproduced by permission of N.T. Pelekanos et al. [47].

Figure 18 shows the PL spectra of three strained MQW [47]. The e_1h_1 lines correspond to an intrinsic free HH-exciton transition, Y lines correspond to the transitions of extrinsic character and sub lines to the substrate contribution. The good optical properties of the samples are demonstrated: (i) by the major contribution of the intrinsic free-exciton recombination, and (ii) by the sharpness of the e_1h_1 lines. These are probably due to the fact that no strain relaxation has occurred in the samples as demonstrated by TEM. Upward white arrows indicate the calculated e_1h_1 position with a model taking into account the strain distribution measured by HRTEM. A good agreement between experimental and theoretical values is obtained.

4.5 Conclusion

Microscopic techniques are quintessential to modern materials science. Through the few selected examples in this chapter one can be convinced that most of the microstructural features which govern the optical quality of optoelectronic materials can be currently determined. In particular, the epitaxial relationships, strain features in heterolayers, dopant and impurity bulk and surface distribution, recombination centers, morphology are exposed. It is hoped that further detailed investigations in materials science will be achieved concurrently with the increasing development and progress of microscopic techniques.

4.6 References

[1] V. Narayanamurti in *Optoelectronic Materials and Device Concepts* (Ed.: M. Razeghi), SPIE Optical Engineering Press, Bellingham-Washington **1991,** p. 1.

[2] John C. C. Fan in *Optoelectronic Materials and Device Concepts* (Ed.: M. Razeghi), SPIE Optical Engineering Press, Bellingham-Washington **1991,** p. 202.

[3] S. M. Sze in *Physics of Semiconductor Devices,* 2nd ed. Wiley & Sons, New York **1981,** Chapter 12, p. 681.

[4] H. Mathieu in *Physique des Semiconducteurs et des Composants Électroniques,* Masson, Paris **1987,** Chapter 9, p. 405.

[5] C. R. M. Grovenor in *Microelectronic Materials* (Ed.: B. Cantos), Adam Hilger Press, Bristol **1989.**

[6] J. Wilson, J. F. B. Hawkes in *Optoelectronics: An introduction* (Ed.: P. J. Dean), Prentice-Hall, Englewood Cliffs, NJ **1983.**

[7] J. W. Edington in *Electron Diffraction in the Electron Microscope,* Philips Technical Library, Eindhoven **1975.**

[8] M. Altarelli in *Les Interfaces et la Liaison Chimique,* 7° Ecole d'Été Méditerranéenne, Toulouse, Les Ed. de Phys., Les Ullis **1988,** p. 221.

[9] G. L. Bir, G. E. Pikus in *Symmetry and Strain-induced Effects in Semiconductors,* Wiley & Sons, New York, **1974.**

[10] F. Capasso, A. Y. Cho, *Surf Sci.* **1994,** *299/300,* 878.

[11] Lester F. Eastman in *Optoelectronic Materials and Device Concepts* (Ed.: M. Razeghi), SPIE Optical Engineering Press, Bellingham-Washington **1991,** p. 41.

[12] P. A. Stadelmann in *Computer Simultion of Electron Microscope Diffraction and Images* (Eds.: W. Krakow, M. O'Keefe), The Minerals, Metals and Materials Society, Pittsburg **1989,** p. 159.

[13] Y. Tomokiyo, S. Matsumura, T. Okuyama, T. Yasunaga, N. Kuwano, K. Oki, *Ultramicroscopy,* **1994,** *54,* 276.

[14] A. F. De Jong, D. V. Dick, *Ultramicroscopy* **1990,** *33,* 269.

[15] A. Ourmazd, D. W. Taylor, J. Cunningham, *Phys. Rev. Lett.* **1989,** *62,* 934.

[16] R. Bierwolf, M. Hohenstein, F. Philip, O. Brandt, G. E. Crook, K. Ploog, *Ultramicroscopy* **1993,** *49,* 273.

[17] J. B. Clegg, in *Growth and Characterization of Semiconductors* (Eds.: L R. A. Stradling, P. C. Klipstein), Adam Higler, Bristol **1990,** p. 87.

[18] A. Huber in *Microcaractérisation des Solides Méthodes d'Observation et d'Analyse* (Ed.: M. Ammou), CRAM, CNRS, Sophia Antipolis, **1989,** Chapter 10, p. 419.

[19] A. M. Huber, G. Morillot, P. Merenda, M. Bonnet, G. Bessonneau, *Appl. Phys. Lett.* **1982,** *41,* 638.

[20] R. R. Olson, P. W. Palmberg, C. T. Hovland, T. E. Brady in *Practical Surface Analysis* (Eds.: D. Briggs, M. P. Seah) Wiley & Sons, New York **1983,** Chapter 6, p. 217.

[21] H. Heral, A. Rocher, *Appl. Phys. Rev.* **1987,** *22,* 201.

[22] P. Trebbia in *Microscopie Electronique en Sciences des Matériaux,* Ecole d'Été du CNRS (Eds.: B. Jouffrey, A. Bourret, C. Colliex), CNRS, 539, Bombannes **1983,** p. 100.

[23] C. Delalande, G. Bastard in *Semiconductor Interfaces: Formation and Properties* (Eds.: G. Le Lay, J. Derrien, N. Boccara), *Springer Proc. Phys.* **1987,** *22,* 340.

[24] H. Lakner, L. J. Balk, E. Kubalek, *Microsc. Microanal. Microstruct.* **1991,** *2,* 293.

[25] S. Gwo, K. J. Chao, C. K. Shih, K. Sadra, B. G. Streetman, *Phys. Rev. Lett.* **1993,** *71,* 1883.

[26] D. B. Holt in *Growth and Characterization of Semiconductors* (Eds.: R. Strading, P. C. Klipstein), Adam Hilger, Bristol **1991,** p. 65

[27] G. Remond, F. Cesbron, R. Chapoulie, D. Ohnenstetter, C. Roques-Carmes, M. Schvoerer, *Scan. Micro.* **1992,** *6,* 23.

[28] J. C. H. Spence in *High-Resolution Transmission Electron Microscopy and Associated Techniques* (Eds.: P. Buseck, J. Cowley, LeRoy Eyring), Oxford University Press, Oxford, **1992,** p. 190.

[29] S. Roberts, *Inst. Phys. Conf. Ser.* **1981,** *60,* 377.

[30] P. M. Petroff, D. V. Lang, J. L. Strudel, R. A. Logan, *Scan Elec. Micr.* **1978,** *1,* 325.

[31] P. M. Petroff, C. Weisbuch, R. Dingle, A. C. Gossard, W. Wiegmann, *Appl. Phys. Lett.* **1981,** *38,* 965.

[32] J. Wang, J. W. Steeds, H. Arnot, *Microsc. Microanal. Microstruct.* **1990,** *1,* 241.

[33] G. Bomchil, A. Halimaoui, I. Sagnes, P. A. Badoz, I. Berbezier, P. Perrett, B. Lambert, G. Vincent, L. Garchery, J. L. Regolini, *Appl. Surf. Sci.* **1993,** *65/66,* 394.

[34] I. Berbezier, A. Halimaoui, *J. Appl. Phys.* **1993,** *74,* .

[35] I. Berbezier, Proceedings of Ecole des Houches, in *Porous Silicon: Science and Technology* (Eds.: J. C. Vial, J. Derrian), Les Editions de Physique and Springer, Les Ullis and Berlin **1995,** p. 207.

[36] A. Albu-Yaron, S. Bastide, D. Bouchet, N. Brun, C. Colliex, C. Lévy-Clément, *J. Phys. I France* **1994,** *4,* 1181.

[37] E. Rosencher, in *Silicon Molecular Beam Epitaxy* (Eds.: E. Kasper, J. C. Bean), CRC Press, Boca Raton **1988,** Vol. 1, p. 161.
[38] S. Luryi and S. M. Sze, in *Silicon Molecular Beam Epitaxy* (Eds.: E. Kasper, J. C. Bean) CRC Press, Boca Raton **1988,** Vol. 1, p. 181.
[39] A. Soufi, G. Brémond, T. Benyattou, G. Guillot, D. Dutartre, I. Berbezier, *J. Vac. Sci. Technol. B* **1992,** *10*, 2002.
[40] J. C. Bean in *Silicon Molecular Beam Epitaxy*, (Eds.: E. Kasper, J. C. Bean) CRC Press, Boca Raton **1988,** Vol. 2, p. 65.
[41] A. Ourmazd, *Mat. Sci. Rep.* **1993,** *9*, 201.
[42] P. Schwander, C. Kisielowski, M. Seibt, F. H. Baunann, Y. Kim, A. Ourmazd, *Phys. Rev. Lett.* **1993,** *71*, 4150.
[43] J. Derrien, J. Chevrier, V. Le Thanh, I. Berbezier, C. Giannini, S. Lagomarsino, M. G. Grimaldi, *Appl. Surf. Sci.* **1993,** *70/71*, 546.
[44] J. Chevrier, P. Stocker, V. Le Thanh, J. M. Gay, J. Derrien, *Europhys. Lett.* **1993,** *22*, 449.
[45] I. Berbezier, J. Chevrier, J. Derrien, *Surf. Sci.* **1994,** *315*, 27.
[46] P. H. Jouneau, A. Tardot, G. Feuillet, H. Mariette, J. Cibert, *J. Appl. Phys.* **1994,** *75*, 7310.
[47] N. T. Pelekanos, P. Peyla, Le Si Dang, H. Mariette, P. H. Jouneau, A. Tardot, N. Magnea, *Phys. Rev. B* **1993,** *48*, 1517.

Handbook of Microscopy
S. Amelinckx, D. van Dyck,J. van Landuyt , G. van Tendeloo

5 Domain Structures in Ferroic Materials

5.1 Ferroic Phase Transitions

A crystal may transform from one crystal structure into another when the thermodynamic environment (temperature, pressure, and chemical composition) is changed. In many cases it is observed that these structural changes are subtle enough so that the crystal is not destroyed during the transformation and that the process is reversible. Much research has concentrated over the last three decades on the investigation of structural phase transitions in which the structural changes maintain the general topology of the crystal and all structural changes occur continuously (or nearly so).

In this case one finds that most phase transitions are correlated with a simple change of the symmetry of the structure: a crystal structure with a certain symmetry described by a space group G_0 (the 'high-symmetry phase') transforms into another structure in which at least one symmetry element of G_0 is missing. The new space group is now called G and the phase is called the 'low-symmetry phase'. Such symmetry conditions curtail significantly the possible atomic movements which lead to the phase transition. In the context of this section, the symmetry conditions are even more severe: we will only consider phase transitions with changes of symmetry elements of rotations, mirror planes, and inversion centers but no changes of any translational symmetries shall occur. Such transitions which maintain translational symmetries are called 'ferroic'. The two most important types of ferroic phase transitions are ferroelastic and ferroelectric ones, which are now discussed exclusively. In fact, most ferroelastic materials show phase transitions in which translational symmetries are also broken (so that the transition is not, strictly speaking, 'ferroic') but still maintain their ferroelastic fingerprint, namely the ferroelastic hysteresis. We will include these materials in the discussion. Details of the symmetry properties and the role of long-ranging strain interactions are given by Salje [1] and Khatchaturyan [2]. We will concentrate here on ferroelastic and ferroelectric phase transitions which are continuous in nature (or nearly so), but not stepwise transitions such as found in martensitic alloys and so forth.

5.2 Ferroelastic Phase Transitions

Ferroelastic materials have an unusual, but technologically most welcome property:

when they are subjected to external stress they first react elastically in a classic manner but when the stress surpasses a certain threshold, the structure suddenly becomes 'soft' and rearranges its atoms in a way which is better adapted to the external stress. This process is reversible. In fact, one finds that there are two or more structural states which are all energy ground states with the possibility to switch the crystal between these states via applied elastic stress. A typical ferroelastic hysteresis is shown in Fig. 1. In this figure the effect of the hysteresis is shown on three different length scales. On an atomistic scale, the relevant part of the crystal structure of $Pb_3(PO_4)_2$ is shown (top). Lead atoms are indicated as black ellipsoids which are surrounded by PO_4 tetrahedra. Around each lead position there are three tetrahedra with their apical oxygens pointing upwards and three tetrahedra with their apical oxygens pointing downwards. The lead position has, thus, topologically a high symmetry coordination with a trigonal space group. At sufficiently low temperatures the energy minimum 'seen' by lead is not on the triad, however, because it is energetically favorable to form some shorter bond distances Pb–O at the expense of elongating others. There are six possibilities to achieve such local distortions, namely by shifting lead positions (in a positive or negative direction) along one of the three channels indicated by arrows. The shifts are very small and can be reversed by external stress, that is, the lead positions can be switched collectively from one such energy minimum to another. Macroscopically, this switching is seen as the ferroelastic hysteresis which relates the angle between two macroscopic directions in the plane shown in the crystal structure, and the external shear stress. The coercive stress is rather small (1.6 bar) and the energy dissipated per switching cycle ($\int e \, d\sigma$, where e is the macroscopic strain and σ is the stress) is much smaller than the excess energy gained by the structural deformation. This observation leads to the question: Why is energy dissipated in a hysteresis much smaller than the total energy needed to displace all atoms via an excited state? The answer lies in the development of the microstructure shown at the bottom of Fig. 1. This figure shows a crystal placed between crossed optical polarizers, so that variations in the optical birefringence are seen as light or dark areas in the crystal. A crystal is shown for three phases of the switching process. Different orientations of the lead displacement appear as black and white areas in this figure. These areas are twins related to twin walls as domain boundaries between them. Only the atoms in these domain boundaries are, at any time, in the excited state between the potential minima of the lead positions of a stress-free crystal. All other atoms are in the equilibrium positions and do not contribute to the energy of the hysteresis. The mechanism by which switching is now achieved is by movement of the domain boundaries. The energy related to the movement of the domain boundary is a virtual energy in the case of lateral movements in an ideal crystal. In real materials, the movement of domain boundaries is hampered by pinning to defects, tilt of domain boundaries, and so forth (Fig. 2). The reversible lateral movement of such a ferroelastic domain boundary (twin wall) can be seen on a fine scale in Fig. 3. In this case, the wall movement is induced by local stresses generated by heating the

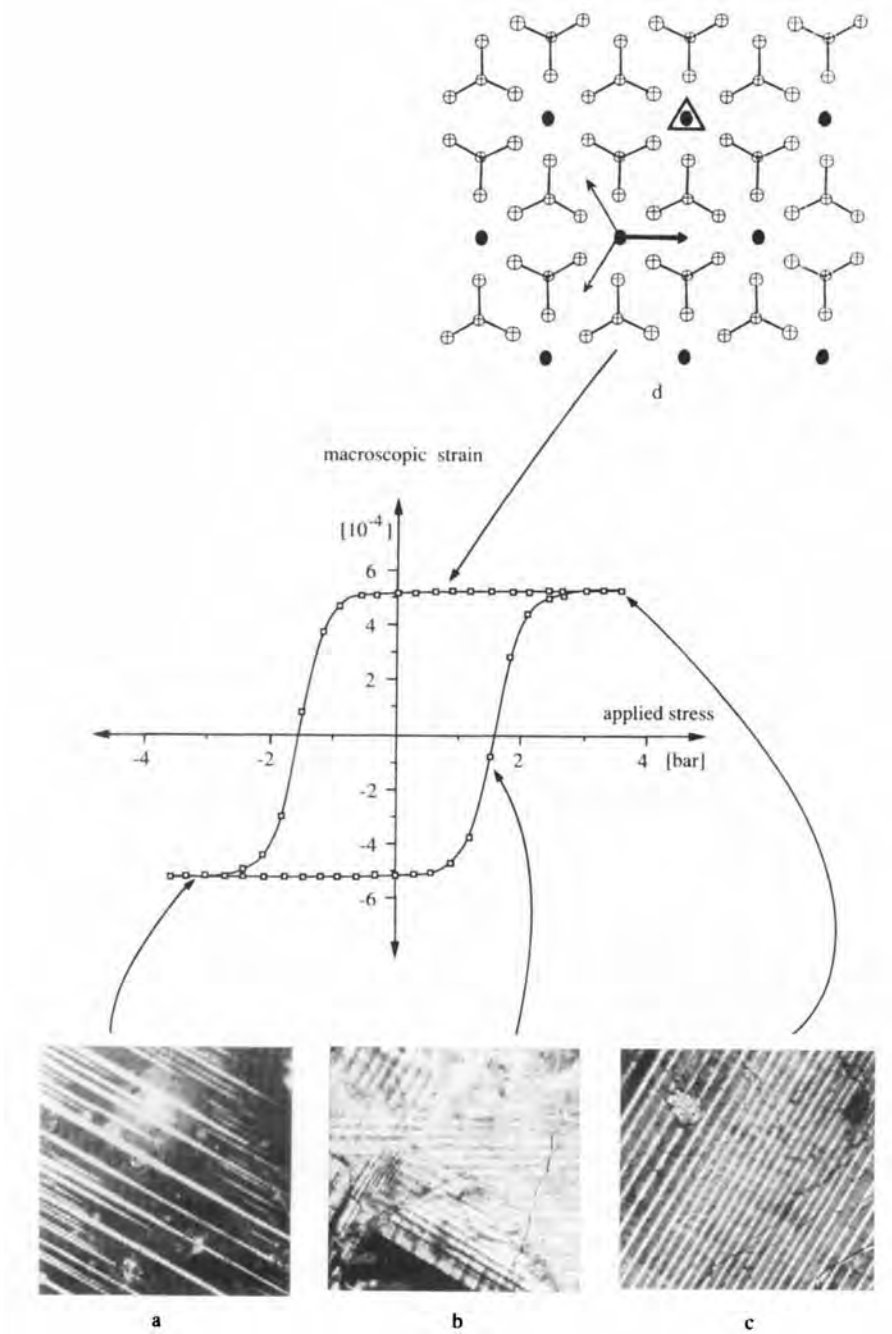


Figure 1. Crystal structure, ferroelastic hysteresis, and optically observed microstructures of $Pb_3(PO_4)_2$ [1]. The black lead atoms are displaced parallel or antiparallel to one of the three axes indicated by errors. The ferroelastic switching involves the reorientation of this displacement, and resulting twin structures are clearly seen under the optical microscope. Twin-free samples can be produced by appropriate external stress.

Figure 2. Observation of twin structures in $Pb_3(PO_4)_2$ with stranded needle domains as observed in the optical microscope using crossed polarizers. The dark and light areas represent the two twin-related domain orientations. The crystal has been slightly sheared, so that the dark domains dominate. The light domains form needle twins with needle tips pinned by defects (stranded needles). Further shear stress will overcome the pinning forces, so that the crystal will become a single domain (i.e., the light domains will vanish completely by retraction and annihilation of the two tips of each needle domain).

sample nonuniformly under the electron beam in a transmission electron microscope. In both ferroelectric and ferroelastic materials the width of the hysteresis, and hence the dissipated switching energy, is mainly determined by the structure and movement of domain boundaries.

Besides the limitation of boundary movements by extrinsic defects, boundary pinning can also occur as a consequence of the interactions between the domain walls. An example is shown in Fig. 4 where a multitude of domains form a complex pattern. The movement of any single domain boundary is hindered by some other domain boundaries. In a more typical situation (Fig. 5), domain boundaries form patterns with two dominant wall orientations. The major interaction is now due to the high strain energy of wall junctions and the consequent repulsion of wall tips. The repelled walls usually form needle domains (comb pattern) which can easily be moved by external stress in the case of most ferroelastic materials.

Before we return to the discussion of the internal structure of domain boundaries, we first introduce the relevant measure for the ferroelastic lattice distortion, namely the spontaneous strain. This strain is generated by the phase transition

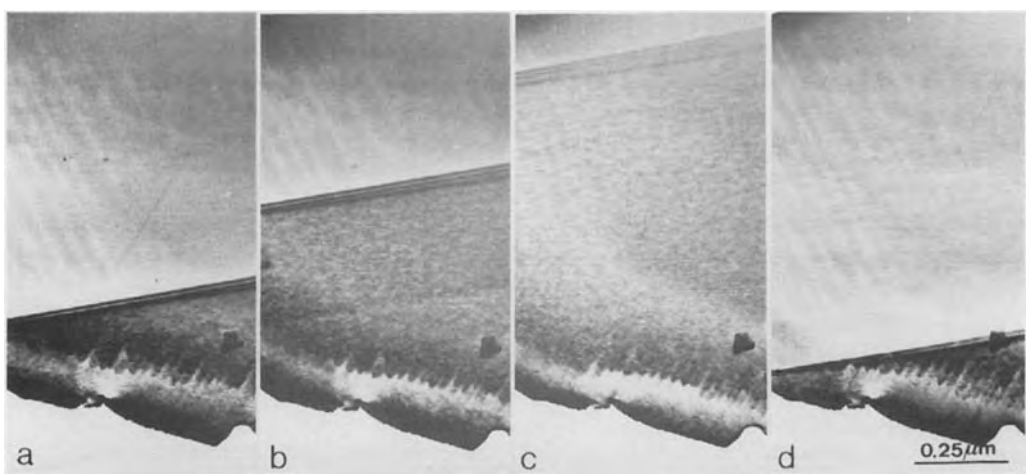


Figure 3. Motion of ferroelastic domain walls in lead vanadate under the influence of beam-induced stress. (Courtesy of EMAT Laboratory, Antwerp; see also Manolikas and Amelinckx [3].)

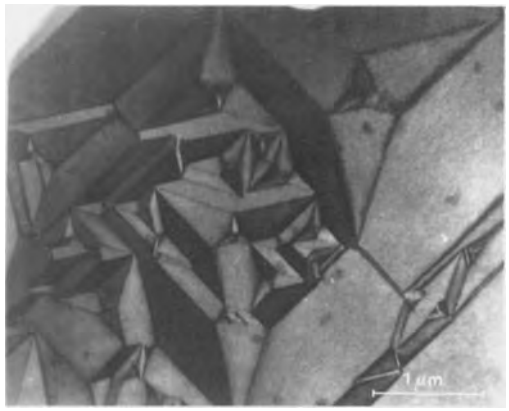


Figure 4. Ferroelastic domain structures with a high number of junctions. Such domain structures can often be aligned in external stress field. (After Manolikas and Amelinckx [3].)

without any external stress applied to the sample.

The ferroelastic strain can be expressed via the size and shape of the crystallographic unit cell (lattice parameters a, b, c, α, β, and γ; reciprocal lattice parameter a^*, b^*, c^*, α^*, β^*, and γ^*) relative to that of the undistorted (i.e., paraelastic) structure. The components of the spontaneous strain are

$$e_{11} = e_1 = \frac{a \sin \gamma}{a_0 \sin \gamma_0} - 1 \tag{1}$$

$$e_{22} = e_2 = \frac{b}{b_0} - 1 \tag{2}$$

$$e_{33} = e_3 = \frac{c \sin \alpha \sin \beta_0^*}{c_0 \sin \alpha_0 \sin \beta_0^*} - 1 \tag{3}$$

$$\begin{aligned} e_{23} &= \frac{1}{2} e_4 \\ &= \frac{1}{2}\left[\frac{c \cos \alpha}{c_0 \sin \alpha_0 \sin \beta_0^*} + \frac{\cos \beta_0^*}{\sin \beta_0^* \sin \gamma_0} \right. \\ &\quad \left. \times \left(\frac{a \cos \gamma}{a_0} - \frac{b \cos \gamma_0}{b_0}\right)\right] \\ &\quad - \frac{b \cos \alpha_0}{b_0 \sin \alpha_0 \sin \beta_0^*} \end{aligned} \tag{4}$$

$$\begin{aligned} e_{13} &= \frac{1}{2} e_5 \\ &= \frac{1}{2}\left(\frac{a \sin \gamma \cos \beta_0^*}{a_0 \sin \gamma_0 \sin \beta_0^*} - \frac{c \sin \alpha \cos \beta^*}{c_0 \sin \alpha_0 \sin \beta_0^*}\right) \end{aligned} \tag{5}$$

$$e_{12} = \frac{1}{2} e_6 = \frac{1}{2}\left(\frac{a \cos \gamma}{a_0 \sin \gamma_0} - \frac{b \cos \gamma_0}{b_0 \sin \gamma_0}\right) \tag{6}$$

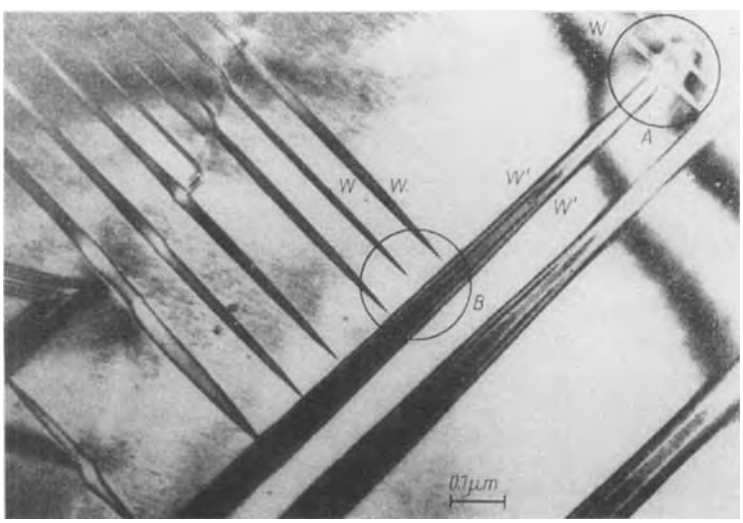


Figure 5. Needle domains with walls W and W′ in $Pb_3(PO_4)_2$. (Courtesy of J. Torres.)

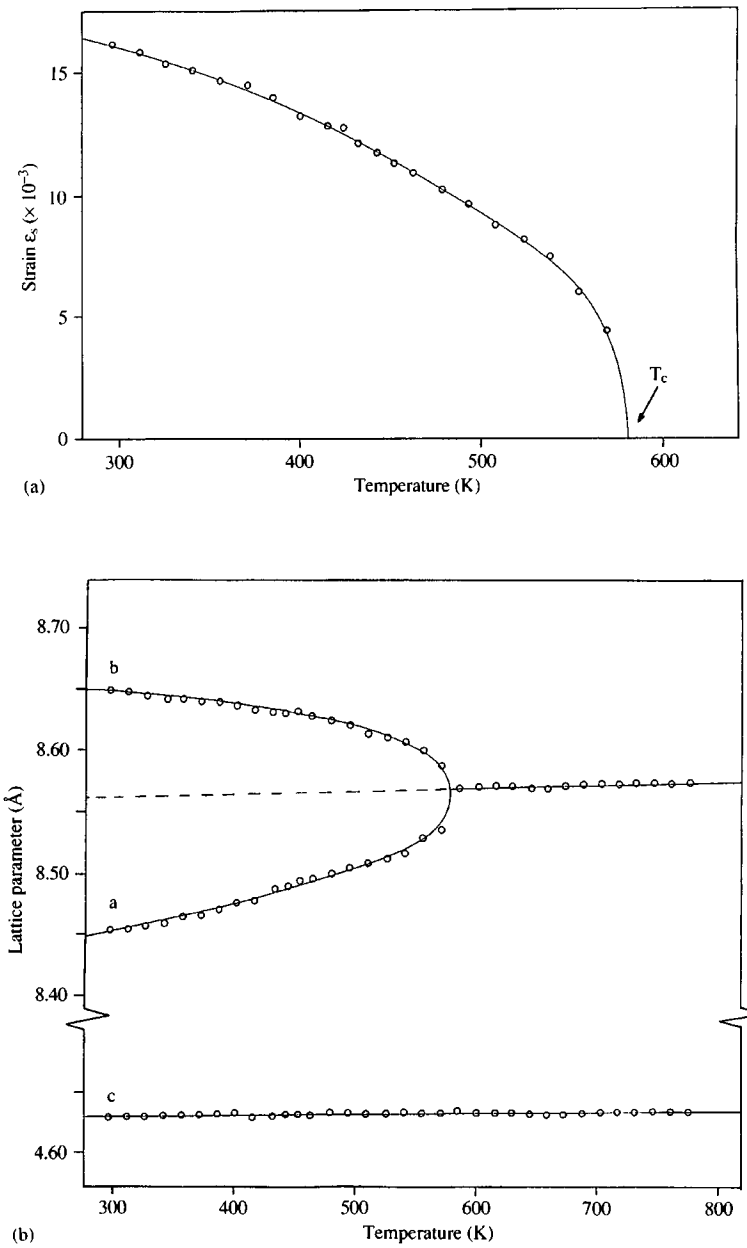


Figure 6. (a) Temperature evolution of the lattice parameters of ferroelastic As_2O_5. The crystal has tetragonal symmetry at $T > T_c$ with $a = b$. At $T < T_c$ the crystal symmetry is orthorhombic. (b) Temperature dependence of the spontaneous strain e_s. The line corresponds to the predicted behavior of a second-order phase transition using Landau theory.

and $e_s = \sqrt{\sum e_i^2}$ is the 'scalar spontaneous strain'. The index '0' indicates the lattice parameter of the high-symmetry phase as extrapolated to the thermodynamic conditions of the low-symmetry phase (for details of the relevant procedures, see Salje [1]). Several of the lattice parameters may assume specific values (e.g. $\alpha = \alpha^* = \gamma = \gamma^* = 90°$ in monoclinic symmetry), so that these equations may

simplify dramatically. A typical example for the temperature evolution of lattice parameters and the scalar spontaneous strain is shown in Fig. 6.

When a ferroelastic crystal is driven through a hysteresis or when the crystal is cooled through the phase transition without the application of external stress, the number of possible domain orientations is predicted by simple symmetry rules. The number N of domain orientations in the low-symmetry form is

$$N = \frac{\text{number of symmetry elements in the high-symmetry phase}}{\text{number of symmetry elements in the low-symmetry phase}} \quad (7)$$

Let us illustrate this rule with the example of the ferroelastic phase transition in NdP_5O_{14} at $T_c = 419\,K$. The high-symmetry form has the point group symmetry *mmm* and the low-symmetry form is ferroelastic with the monoclinic point group $2/m$. The total number of elements in *mmm* is eight, the total number of elements in $2/m$ is four, so that $N = 8/4 = 2$, that is, there are only two different domains and, consequently, two types of wall between them.

Let us now ask what the orientations of twin walls between ferroelastic domains are. If we assume that in some approximation the volume of the crystal is not changed during the phase transition, we find the simple condition of stress free walls:

$$\sum_{i,k=1,2,3} (e_{ik}(S_1) - e_{ik}(S_2)) X_i X_k = 0 \quad (8)$$

where S_1 and S_2 denote the two domains 1 and 2 and X_i are coordinates on the orthonormal axes of the spontaneous strain (e.g. $X_1 = X$, $X_2 = Y$, $X_3 = Z$). As an example, consider the transition $2/m \rightarrow \bar{1}$ with

$$e_{ik}(S) = \begin{pmatrix} 0 & e_{12} & 0 \\ e_{12} & 0 & e_{23} \\ 0 & e_{23} & 0 \end{pmatrix} e_{ik}(S_2)$$
$$= -e_{ik}(S_1) \quad (9)$$

which leads to the solution for the walls $Y = 0$ (i.e., the (010) plane) and $Z = (e_{12}/e_{23})X$ which is a plane perpendicular to (010). Twin walls which are fully determined by symmetry are called W walls (here (010)), and other twin walls are denoted W′. For a pioneering study of ferroelastic structures see, for example, Manolikas and Amelinckx [3].

5.3 Wall Energies and Wall Profiles

Although twin walls are rather thin on a macroscopic length scale, they can be described approximately in continuum theory. In fact, it is a misconception to assume that domain boundaries are thin on a level of atomic diameters. Typical wall structures can be seen in Fig. 7 with (a) a 'wide' domain boundary with $w \approx 7\,nm$ and (b) a 'thin' domain boundary with $w \approx 0.7\,nm$. Both domain boundaries occur typically in ferroelastic $YBa_2Cu_3O_{7-\delta}$ where wide boundaries are often related to oxygen deficit or impurity atoms such as cobalt or aluminum replacing copper [1, 5]. Wide domain boundaries also occur for antiphase boundaries such as in $Pb_3(VO_4)_2$ (Fig. 8) or titanite $CaTiSiO_4$. In the example of Fig. 8 the width w of the antiphase boundary is $\sim 4\,nm$. The most commonly adopted

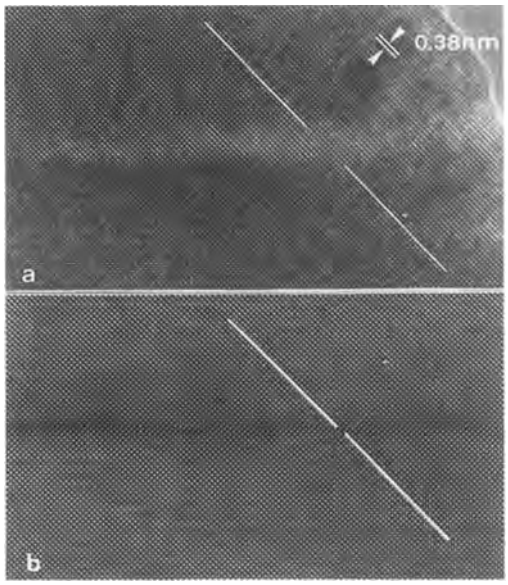


Figure 7. High-resolution transmission electron microscope image of (a) wide and (b) narrow twin boundaries in $YBa_2CU_3O_{7-\delta}$. (Courtesy of EMAT Laboratories, Antwerp.)

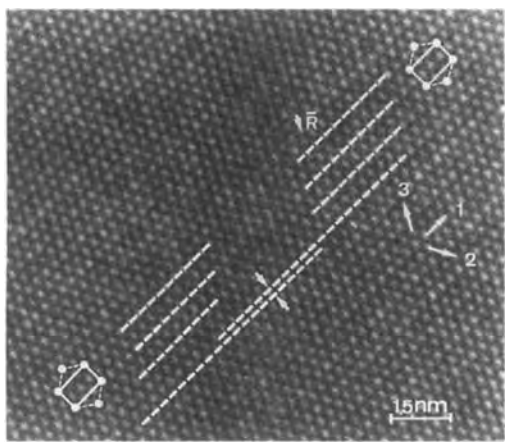


Figure 8. Wide antiphase boundary high-resolution electron microscope view in lead vanadate. Note the gradual shift of lattice rows of types 1 and 2. (Courtesy of EMAT Laboratories, Antwerp.)

approach to analyze domain boundaries theoretically stems from the analysis of displacive phase transitions using an expression for the macroscopic excess Gibbs free energy of the low-symmetry phase with respect to the high-symmetry phase:

$$\Delta G = \tfrac{1}{2}A(T - T_c)Q^2 + \tfrac{1}{4}BQ^4 + \cdots + \tfrac{1}{2}g(\nabla Q)^2 \qquad (10)$$

where we generalize the strain e_{ik} to the relevant thermodynamic quantity which is the order parameter Q. The first part of this Landau–Ginzburg potential describes the thermodynamic behavior of the uniform crystal, and the last term is the energy related to variations of the order parameter in space (∇ is the gradient operator $\partial/\partial r$). The equilibrium condition $\partial G/\partial Q$ has besides the uniform solution in the ferroelastic phase $Q = Q_0 = \sqrt{(A/B)(T_c - T)}$ at $(T < T_c)$ also the solution for a single twin wall,

$$Q = Q_0 \tanh \frac{X}{w} \qquad (11)$$

where

$$w = \frac{1}{2}\left(\frac{g}{A(T_c - T)}\right)^{1/2} \qquad (12)$$

is the thickness of the twin wall (e.g., the width of the strain field around the wall). The energy E_w of the wall is directly proportional to its thickness, so that the fundamental physical parameter which determines both w and E_w at sufficiently low temperatures is the ratio g/AT_c. In a first conjecture, it was commonly assumed that g/AT_c is equal to d^2 where d is the lattice repetition length normal to the twin wall. More careful experimental work using X-ray diffraction analysis has revealed that g/AT_c may be substantially greater, with resulting wall thicknesses in framework structures ranging from 0.5 to 10 nm [4, 5]. As the precise knowledge of

g/AT_c is central to the theoretical treatment of domain structure, the experimental determination of the thickness of twin wall is tantamount for progress in this field of research.

5.4 The Internal Structure of Twin Walls

Ferroelastic and ferroelectric framework structures often show a multitude of phase transitions. In fact, it meant a breakthrough for the quantitative analysis of ferroelastic transitions when it was discovered that virtually all phase transitions (i.e., their relevant order parameters Q_i) influence each other even when the actual transition temperatures are far apart. This mutual interdependence is analyzed via energies which represent the coupling between various order parameters. In a most simple expansion of the energy expression, we find

$$\begin{aligned} G = {} & \tfrac{1}{2}A_1(T - T_c^1)Q_1^2 + \tfrac{1}{4}B_1Q_1^4 \\ & + \tfrac{1}{2}g_1(\nabla Q_1)^2 + \tfrac{1}{2}A_2(T - T_c^2)Q_2^2 \\ & + \tfrac{1}{4}B_2Q_2^4 + \tfrac{1}{2}g_2(\nabla Q_2)^2 \\ & + \lambda_1 Q_1 Q_2 + \lambda_2 Q_1^2 Q_2^2 \end{aligned} \tag{13}$$

There are strong symmetry constraints for $\lambda_1 \neq 0$ but none for $\lambda_2 \neq 0$. What is the internal structure of a domain wall of Q_1 for $\lambda_1 = 0$ and $\lambda_2 \neq 0$? It is obvious that for sufficiently weak coupling the equilibrium conditions $\partial G/\partial Q_1 = 0$ will lead to practically the same wall profile as before. This situation is rather unphysical, however, and cases with strong coupling may well be more realistic. In this case, wall profiles are significantly changed. The reason is that $\lambda_2 > 0$ means that the order parameters tend to suppress each other, that is, if a structure is distorted according to Q_1 it will not allow much distortion with respect to Q_2 and vice versa. A special situation occurs in the wall: Q_1 becomes zero so that the suppression vanishes and Q_2 can appear. This effect can be visualized if the wall profiles are plotted in the vector space of the two order parameters (Fig. 9). The bulk of the material is dominated by Q_1, and Q_2 appears in the wall. There are two solutions, however. In the first solution Q_2 is maximal in the middle of the wall with either $+Q_2$ or $-Q_2$ at the center. The state parameter is then a vector (Q_1, Q_2) which rotates positively or negatively. The wall now has a chirality + or −. In the second solution, Q_2 appears at the edges of the wall but also disappears at the center. In this case the vector (Q_1, Q_2) rotates twice, and the wall is called bichiral. Theoretical work has predicted a multitude of novel phenomena related to structural states inside twin walls which have no equivalent state in the bulk of the crystal [6].

5.5 Bending of Domain Walls and the Compatibility Relationship

The orientation of ferroelastic twin walls (W, W′) is determined by the condition that adjacent domains can match coherently. Any deviation from this orientation could, in principle, result in the creation of topological defects, and the crystal structure would cease to be coherent. In

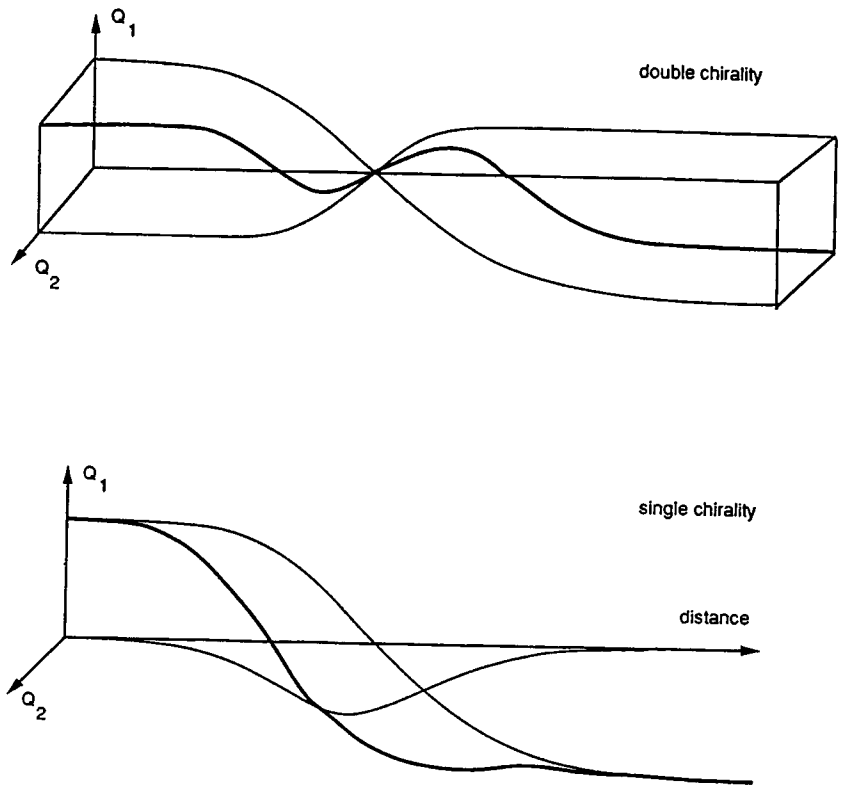


Figure 9. Wall profiles for twin domains with two coupled order parameters ($\lambda_2 Q_1^2 Q_2^2$). The relevant order parameter Q is plotted (thick line) as a vector sum of Q_1 and Q_2. The interaction between the two order parameters may lead to single or double rotation of Q, so that the internal wall structure becomes chiral with respect to the relationship between Q_1 and Q_3.

framework structures with sufficiently large unit cells this does not, in general, happen because the required energy to form dislocations is too large. As a result, the defect is compensated by internal adjustments of the atomic positions inside the unit cell. The penalty for such adjustments is paid for in terms of lattice strain, which is not identical with the spontaneous strain. Let us consider wiggled walls and wall corners, which are commonly observed in ferroelastic materials (Fig. 10). The quantitative description of the strain fields is based on the compatibility relationship, which allows us to calculate the strain components necessary to maintain the coherency of the lattice when a twin wall bends. If no dislocation is generated, the displacement of any atomic position u_i has to be continuous with respect to spatial variations up to second order. In this case the compatibility conditions mean that mixed second spatial derivatives of the displacement field cannot depend on the order in which the derivatives are taken:

$$\frac{\partial^2 u_i}{\partial x_j \partial x_k} = \frac{\partial^2 u_i}{\partial x_k \partial x_j} (j \neq k) \tag{14}$$

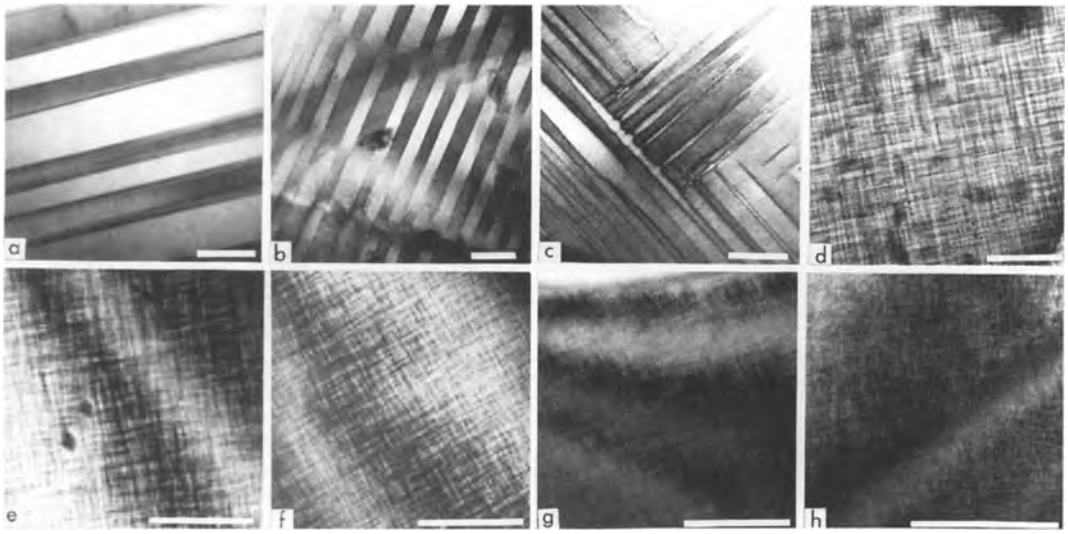


Figure 10. Characteristic microstructures in ferroelastic $YBa_2Cu_{3(1-x)}Co_{3x}O_7$ with $x = 0$ (a), 0.01 (b), 0.02 (c), 0.025 (d), 0.028 (e), 0.03 (f), 0.05 (g), and 0.07 (f). The scale bar is 0.1 μm. Twinning occurs for (a)–(c), and a tweed pattern is seen for all other compositions. Wall junctions, needle domains, and "wiggled' walls are seen in (c). (Courtesy of A. Putnis, Cambridge.)

This condition has substantial consequences when reformulated in terms of strain components e_{ik} rather than the displacements u_i. Let us consider a case in which we consider a trace of a twin wall in the xy plane, assuming that the wall does not bend in the z direction (e.g., the direction of the electron beam). In this case the compatibility relation becomes

$$\frac{\partial^2}{\partial x^2} e_{yy} + \frac{\partial^2}{\partial y^2} e_{xx} = 2 \frac{\partial^2}{\partial x \, \partial y} e_{xy} \qquad (15)$$

Let us now consider a spontaneous strain e_{xy}. For a planar wall along the xy direction (e.g., the (110) plane) the derivative $\partial^2 e_{xy}/\partial x \, \partial y$ vanishes, and there are no strain components e_{xx} or e_{yy}. As soon as the wall wiggles, the second derivative $\partial^2/\partial x \, \partial y$ measures the curvature of the wall. In order to fulfill the compatibility relation (Eq. (15)), spatial variations of e_{xx} and e_{yy} must exist, that is, these strain components can no longer vanish. This simple result implies that walls generated by the primary (i.e., spontaneous) strain will automatically generate secondary strain fields if they bend. These secondary strains are, indeed, the most sensitive indicators of wall deformations and largely determine their orientational fluctuations. A typical example is shown in Fig. 11, where pairs of domain walls join in order to form needle domain close to orthogonal domain walls. This image was taken in diffraction contrast so that the strain fields become observable. No or little strain variation is seen near the part of the domain structure with straight domain walls (top and bottom). Increasing strain modulation appears as contrast inside the needle domains and between the needles related to wall bending. Maximum strain contrast is seen at the needle tips. This strong contrast is similar to (but not

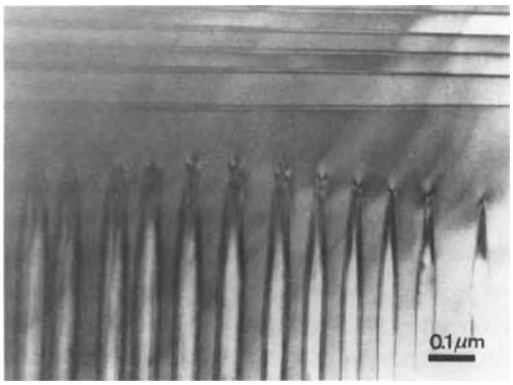


Figure 11. Comb pattern of needle domains in $YBa_2Cu_3O_{7-\delta}$, seen in diffraction contrast. Large strain fields can be seen near the tip of the needles. (Courtesy of EMAT Laboratories, Antwerp.)

necessarily generated by) dislocations at the needle tips. The similarly between strain fields from curved domain walls (without topological defects) and those generated by dislocations has led to the notion of 'pseudo-dislocations' for the theoretical description of strain fields [1]. Care should be taken not to confuse actual dislocations and 'pseudo-dislocations', which are only an elegant mathematical trick for the evaluation of the relevant strain fields of bent domain walls in terms of an equivalent behavior of hypothetical walls with in-built dislocations.

5.6 Wall–Wall Interactions in Ferroelastic Materials

Ferroelastic twin walls interact weakly for parallel walls at large distances $L \gg w$ (with predominantly attractive forces when surface effects are ignored). Strong repulsion occurs for distances of a few w, and it appears that interwall spacings smaller than some $5w$ have never been observed experimentally. Another predominant interaction occurs between walls which may form intersections. Typical examples are W and W′ walls with angles of $\pi/2 \pm \omega$ between each other. The angle ω measures the spontaneous strain. This situation is illustrated in Fig. 12. Here two walls are aligned along planes which were orthogonal in the high-symmetry phase but are sheared against each other in the low-symmetry phase. If we now measure the angle around the junction, we find that this angle is either $2\pi + \omega$ or $2\pi - \omega$, depending on the orientation of the twin planes. The compatibility condition now requires that the crystal is macroscopically sheared in order to compensate for the over- or unoccupied wedges near the junction.

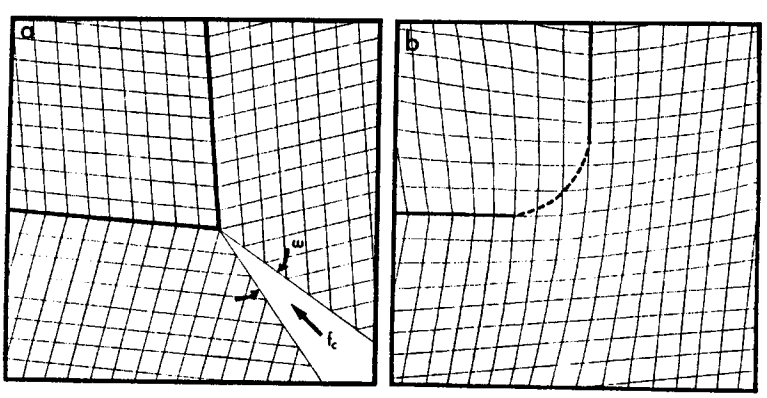


Figure 12. Junction formation between two ferroelastic walls. (a) The angular misfit ω between two parts of unstressed lattice adjacent to the corner of two twin domains causes an effective force, f_c, which (b) closes the open wedge. This closure causes the domain walls to curve, and generates elastic stresses which emanate from the corner.

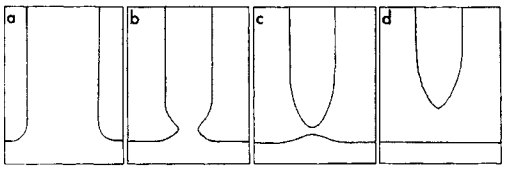


Figure 13. Formation of needle twins. (a) Two right-angled domains attract each other, and (b) bend the domain walls. The area of the grey domains increase with respect to the white domain. (c) When the two junctions join, the grey domains connect whereas the white domain is split into two parts. (d) The new needle domain will not retract under the influence of the interaction between the needle tip and the planar wall, also reducing the wall energy of the needle domain.

This energy, modified by some adjustment of the wall positions, is the characteristic excess energy of the junction. Thermal annealing of ferroelastic microstructures is largely related to the annihilation of such high-energy junctions.

The strong stress fields emitted by such junctions often lead to S-shaped wiggles of neighboring walls with the maximum deformation just above and below the junction.

The most common microstructures contain needle domains as shown in Fig. 5. The most common origin of such needle domains relates to their proximity to another, perpendicular twin wall. Figure 13 shows how a needle forms out of two junctions which attract each other. Once the needle is formed, it may retract and thereby reduce the total surface area of the twin walls. This retraction can be stopped by defects so that annealed samples often show such 'stranded' needle domains (see Fig. 2). The pinning force of defects is sometimes small enough to be overcome by mechanical tapping of the crystal or by local heating with the electron beam in the electron microscope. Once the needle domain is unpinned it will (under open-boundary conditions) continue to retract.

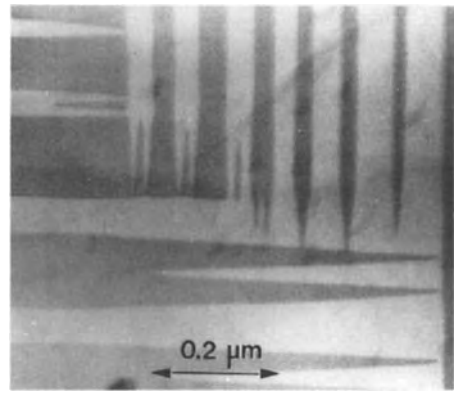


Figure 14. Needle domains with some forked domains in $YBa_2Cu_3O_{7-\delta}$. Forks appear for wide domains whereas simple needles occur for thin domains. (Courtesy of EMAT Laboratories, Antwerp.)

Typical needle domains with pairs of needles forming forks are shown in Fig. 14. The origin of these forks is similar to that of simple needles. The difference stems from the greater width of the forked domains. This larger distance between domain boundaries requires a stronger curvature of the walls in order to form needles, so that it is energetically advantageous to split the needle into two needles (vertical black, pointing downwards in Fig. 14), separated by a third needle between them (white, pointing upwards in Fig. 14). Such patterns do not occur for smaller distances between walls as seen for the horizontal needles at the bottom of the right-hand side in Fig. 14.

5.7 Tweed Microstructures as the Result of Kinetic Processes at $T < T_c$

Rapid quench of ferroelastic material (in particular those involving atomic ordering

processes) may result in a characteristic tweed microstructure (see Fig. 10d–h). Such kinetic microstructures are also observed in materials in which no such structures are seen in thermodynamic equilibrium (e.g., Na-feldspar). Similar tweed structures are sometimes claimed to exist in thermodynamic equilibrium at $T > T_c$, although the experimental evidence appears to be controversial in defect-free materials. It is generally accepted, however, that thermodynamic fluctuations at $T \gtrsim T_c$ are always highly anisotropic, so that each snapshot on a short enough timescale will look tweed-like, and it is conceivable that rapid quench will, indeed, maintain 'equilibrium' tweed at low temperatures.

Such initial tweed may somewhat coarsen kinetically at $T < T_c$ and form a higher contrast tweed with patchy areas separated by wall-like features. Such kinetic tweed is encountered in most ferroelastic materials of the order/disorder type. Extensive computer simulation of tweed patterns has shown that tweed develops into rather dense twin patterns (so-called stripe patterns) via the formation and retraction of needle-shaped domains (Fig. 15).

The term 'tweed' is used here in a generic sense indicating a characteristic microstructure of an interwoven pattern not dissimilar to those seen in woven tissues. Tweed shows usually some preferential orientation of two orthogonal 'weaving' directions.

Although the general appearance of tweed structures is very characteristic, detailed studies show large variations between tweed patterns in different materials and of different physical origins. A tweed pattern with strong contrast on dark field observation is sometimes related to exsolution phenomena and is unrelated to structural phase transitions altogether. Another type of tweed, usually seen as a weak modulation of the strain contrast, exists for all thermodynamic fluctuations which are coupled to structural strain. This structural strain does not usually develop isotropically but depends strongly on the orientation of the fluctuation. Fluctuation waves will often be the largest along two

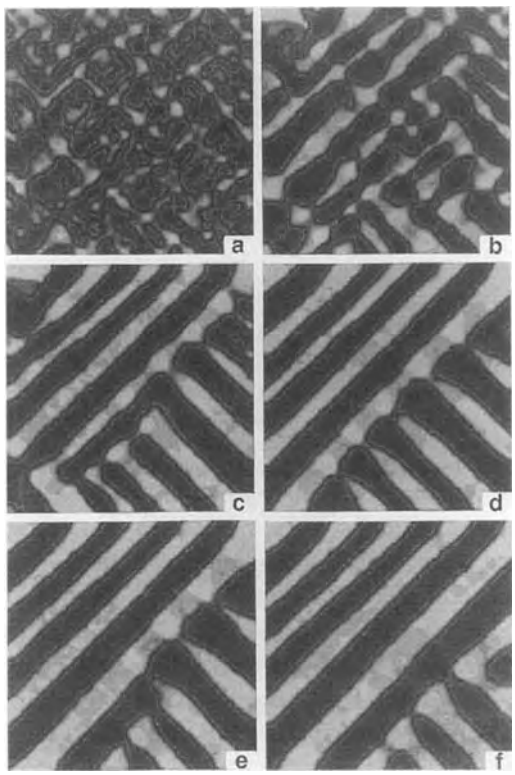


Figure 15. Maps of strain order parameter obtained during annealing at temperature $T_a = 0.43T_c$, after quenching from $4T_c$. (a)–(f) Maps corresponding to annealing times t = 10, 50, 200, 1000, 2000 and 4000 Monte Carlo steps per site. Note the immediate appearance of tweed in (a) due to the pre-existence of embryos even at $2T_c$, the sharpening up of microdomain walls in (b), the coarsening in (b) and (c), and the formation of needle domains in (d)–(f).

elastically 'soft" directions. These soft directions can easily be derived if the elastic properties of the material are known [7].

Fluctuation-induced tweed, in contrast to other forms of tweed, need not display a characteristic length scale of the tweed pattern. Simple fluctuations are described in the Ornstein–Zernicke limit with a Gaussian distribution of the wavevectors ($\bar{k} = (2\pi/l) \cdot \hat{e}$, where l is the length of a fluctuation and $\hat{e}$ is its orientation) centered around $k = 0$. This theoretical result is simply based on the fact that long-wavelength fluctuations are favored because they consume less of the gradient energy in Eq. (10).

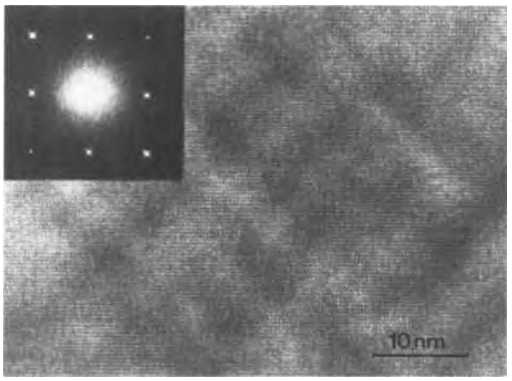


Figure 16. High-resolution image of the tweed pattern in $YBa_2Cu_3O_{7-\delta}$ and the diffraction pattern. The pattern shows preferential elongations in the ⟨110⟩ directions and weakly developed periodicities. (Courtesy of EMAT Laboratories, Antwerp.)

Although fluctuation tweed is elegantly described in such terms, and some experimental evidence from diffraction experiments is compatible with these ideas [8], there is considerable doubt that they describe the physical situation correctly at temperatures other than at $T \gg T_c$. At lower temperatures two additional features appear. Firstly, the amplitude of the fluctuations increases so that energy terms which are anharmonic in the fluctuation amplitude become important. This makes the description of fluctuations in terms of harmonic waves inadequate and requires the consideration of solitary excitations. A simple physical picture is that the areas of positive and negative fluctuations form 'patches' separated by wall-like discontinuities. The same 'patchy' tweed structure also follows from higher-order gradient energies which favor periodic, nonsinusoidal microstructures. Any of these effects will introduce a characteristic length scale for fluctuation tweed, namely either the thickness of wall-like features or a preferential periodicity of the elastic deformation wave. Simple numerical estimates lead for both cases to periodicities of the tweed pattern of 5–15 nm. The challenge for the experimentalist is to determine if such characteristic length scales exist at $T > T_c$. One of the probably most advanced observations in this field is shown in Fig. 16, which shows a patchy tweed pattern with wide range of wavelengths and a characteristic length of $\lambda_m \approx 10$ nm.

Kinetic tweed patterns at $T < T_c$ are better understood than fluctuation tweed at $T > T_c$. The metastable microstructure with the longest lifetime is the twin structure. If a crystal with fluctuation tweed at $T > T_c$ is quenched to $T < T_c$, the tweed structure will rapidly 'square-up', that is, it will form patches where the order parameter inside each patch is close to its equilibrium value. Walls between patches are now very similar to twin walls with a high number of junctions between orthogonal walls. As these junctions generate large strain energies they will be gradually eliminated, and the tweed structure

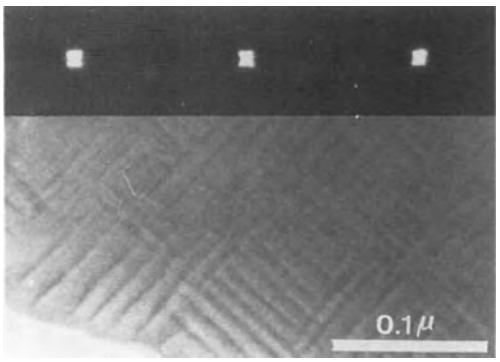


Figure 17. Tartan pattern (coarse tweed) and diffraction pattern of rapidly cooled $YBa_2Cu_3O_{7-\delta}$. (Courtesy of EMAT Laboratories, Antwerp.)

coarsens (Fig. 17). Such coarse tweed structures are sometimes called 'tartan' structures. Further coarsening of tartan to a stripe pattern occurs via the formation and retraction of needle domains (see Fig. 5). A typical sequence of tartan and stripe patterns under beam heating in the electron microscope is shown in Fig. 18.

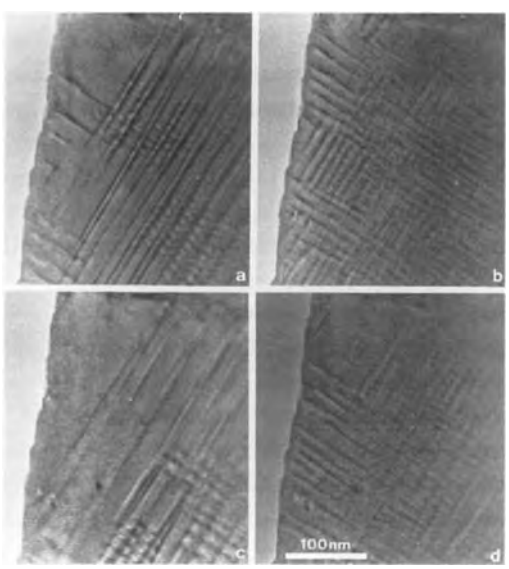


Figure 18. Time evolution of the tweed (tartan) pattern in $YBa_2Cu_3O_{7-\delta}$ at room temperature. (Courtesy of EMAT Laboratories, Antwerp.)

5.8 Ferroelectric Domains

Electron microscopic work on a ferroelectric wall is hampered by the difficulty of making the reversion of the ferroelectric polarization visible. Since the mid-1960s, Takagi and Suzuki [9] and others have applied electron microscopy decoration techniques to study domain structures in ferroelectric crystals, although the exact physical mechanism of the AgCl–C coating revealing polarization orientation in replicas is still not fully understood. Undeniably, the technique succeeds in showing clear polarization reversal and also seems to allow first attempts to measure the thickness of ferroelectric domain boundaries. Hilczer et al. [10] found that the image of a TGS-cleaved surface shows a domain wall of some 12 nm. Direct observations on ferroelectrics pose unusual problems, since specimens tend to be unstable under the electron beam, and there are difficulties correcting astigmatism. Despite these difficulties, some direct observations in $LiTaO_3$ seem to show very thin domain walls ($\omega \lesssim 0.28$ nm). This observation [11] relates to the observation of the coupled ferroelastic strain, however, and may not directly reflect the domain profile of the spontaneous polarization. Measurements of strain profiles are equally difficult on this length scale, so that the lower limit of reliable information on the value of w in ferroelastics may still be some 0.5 nm. Wider walls, extending over some 5 nm, have also been identified in $KNbO_3$ by Bursill and Lin [11].

5.9 References

[1] E. K. H. Salje, *Phase Transitions in Ferroelastic and Co-elastic Crystals*, Student ed., Cambridge University Press, Cambridge **1993.**

[2] A. G. Khatchaturyan, *Theory of Structural Transformations in Solids*, Wiley, New York **1983.**

[3] C. Manolikas, S. Amelinckx, *Phys. Status Solidi* **1980,** *A60*, 607.

[4] B. Wruck, E. K. H. Salje, M. Zhang, T. Abraham, U. Bismayer, *Phase Trans.* **1994,** *48*, 135.

[5] J. Chrosch, E. K. H. Salje, *Physica C* **1994,** *225*, 111.

[6] B. Houchmanzadeh, J. Lajzerowicz, E. Salje, *J. Phys. C: Cond. Matter* **1991,** *3*, 5163.

[7] I. Tsatskis, D. A. Vul, E. K. H. Salje, V. Heine, *Phase Transitions* **1994,** *52*, 95.

[8] D. DeFontaine, L. T. Wille, S. C. Moss, *Phys. Rev. B* **1987,** *36*, 5709.

[9] M. Takagi, S. Suzuki, *J. Phys. Soc. Jpn* **1967,** *23*, 134.

[10] B. Hilczer, L. Szczesniak, K. P. Meyer, *Ferroelectrics* **1989,** *97*, 59.

[11] L. A. Bursill, P. J. Lin, *Ferroelectrics* **1986,** *70*, 191.

Handbook of Microscopy
S. Amelinckx, D. van Dyck,J. van Landuyt , G. van Tendeloo

6 Microscopy of Structural Ceramics

6.1 Introduction

Ceramics play an ever more important role in science as well as in applications. Ceramics can be divided into different groups, that is, whiteware ceramics, structural ceramics, and electroceramics [1]. Whiteware ceramics deals mainly with the processing of clay and other minerals into objects for everyday life. Electroceramics are oxides, carbides, or nitrides of metal alloys which are used because of their specific magnetic or electrical properties. In contrast, advanced structural ceramics are mainly used in specific components working under extreme conditions such as high temperature, high stress, or high friction. Structural ceramics can be either oxides, carbides, or nitrides of different cations, such as aluminum, silicon, or boron.

At ambient temperatures most ceramics are brittle; they do not show any plastic behavior upon loading. Therefore the mechanical properties of these ceramics are mainly determined by the defects present in them. It is, therefore, essential to characterize the microstructure and defects of these materials. This microstructural characterization includes all the dimensions from the atomic scale to the component (Fig. 1). For example, silicon nitride (Si_3N_4) parts are now used as engine components, e.g., valves, turbo chargers. The Si_3N_4 engine parts have a complex shape and to densify and process the material without flaws is rather complicated. Flaws and defects have to be identified and studied, and the microstructure has to be investigated—often with resolutions from optical microscopy to the atomic level. For example, large defects such as voids or inclusions can easily be identified by optical microscopy. The grain size and grain size distribution of the different materials can be determined by optical microscopy [2] and by scanning electron microscopy (SEM) [3], and specific details of the microstructure can be investigated by scanning tunneling microscopy (STM) or atomic force microscopy (AFM) [4, 5]. Quantitative results on the microstructure can be elaborated using standard quantitative metallography techniques [2, 6]. Structural materials are usually polycrystalline. They consist of many grains in different orientations. The grain boundaries and interfaces of these materials are very important defects, often controlling the properties. Therefore it is essential to investigate the structure and chemical composition of grain boundaries; this requires microscopy at the ultimate resolution of transmission electron microscopy [7, 8].

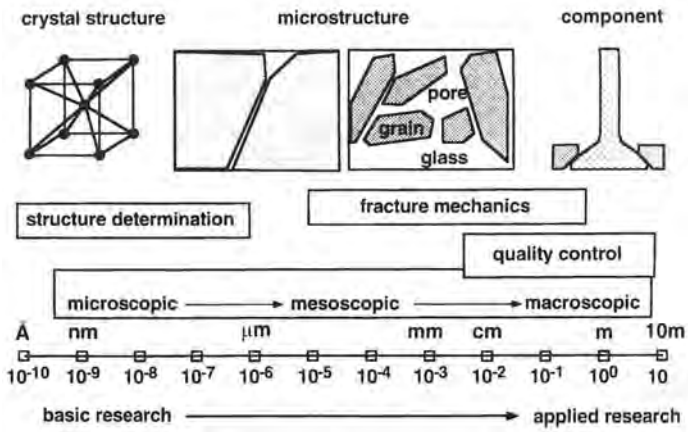


Figure 1. Microstructure ranging from crystal structure to engine components.

Several conferences on the microstructure of ceramic materials have been held over the last few decades. The proceedings of these conferences on ceramic microstructures have been published and reflect the importance of understanding the microstructure of these materials [9–12].

In this chapter, the different techniques for analyzing the microstructure of ceramics will initially be summarized. The important features of the microstructure of three prominent ceramics will be described: alumina, zirconia, and silicon nitride. Finally, some remarks on the microstructure of composites will conclude the chapter.

6.2 Different Imaging Techniques

6.2.1 Remarks on Specimen Preparation

Most ceramics are nontransparent, therefore the optical microscope is used in the reflection mode. To make different grains and inhomogeneity visible, the specimen has to be etched [2, 13]. This involves a sequence of mechanical grinding and polishing. Special materials are required for each ceramic in order to achieve the best possible polishing. Similar to metallographic preparation, the following steps have to be carried out to prepare ceramics for microstructural investigations. Firstly, the proper tool has to be selected for grinding, polishing, cleaning, and etching. However, owing to the brittleness, porosity, and chemical resistance of ceramics, they are quite often difficult to polish compared to metals. Polishing is usually carried out using diamond paste, followed by small silicate particles. If the specimen is polished too much, then the region beneath the polished surface will be damaged leading to artifacts in the microstructure. The specimen has to be etched after polishing. This makes the different grains and components visible. However, the etching of ceramics requires very aggressive chemicals which are often very difficult to handle. Therefore specimens are often etched using plasmas. This etching involves ion etching followed by bombardment with electrons. Thermal

etching is an important technique for revealing the ceramic microstructure. For this technique a polished specimen is annealed in a high-temperature furnace. The temperature and time have to be chosen in such a way that the microstructure is not modified (e.g., by grain growth). Microstructural characteristics can be made visible by thermal grooving [1].

6.2.2 Optical Microscopy

Classic metallographic techniques can also be applied to ceramics. The grain size and different phases can be identified if the grain size is large compared to the resolution of the microscope. The addition of interference layers on top of the specimen improves the phase contrast. This cannot be done as successfully for ceramics as for metals, since ceramics have a much smaller coefficient of reflection compared to metals; also some ceramics are transparent [2, 13]. At the interface between the interference layer and the ceramic the conditions for multiple reflection leading to the desired interference are not fulfilled. Ceramic microstructures can only be studied by optical interference microscopy under the following conditions: (i) If the microstructure consists of different phases and the phases are opaque. The interference layers on the ceramics are mostly ZnSe, ZnTe, and ZnS. (ii) If amplification by multiple reflection is obtained due to the interference layer possessing a reflection index smaller than that of the ceramics. The best material is kryolite (Na_3AlF_6). (iii) If layered specimens and metal/ceramic composites are contrasted by reactive sputtering [2, 13]. The difference in reflectivity between the different components can be increased or reduced depending on the sputtering material.

6.2.3 Microstructural Studies with the Scanning Electron Microscope (SEM)

The resolution of optical microscopy is limited essentially by the wavelength of light. A better resolution can be achieved by using electrons as a 'light source'. Depending on the accelerating voltage, the resolution can be improved [3] by using a scanning electron microscope (SEM). A small probe is scanned over the specimen surface. Ceramics are often electrically insulating materials. Therefore they have to be coated with a conducting layer of carbon or gold [3]. Recently, SEMs have become available that operate at low voltages ($\lesssim 1$ keV) and with a field emission source. The resolution is impressively improved and details such as atomic steps at surfaces can be seen. Operating at low voltages makes coating with a conducting layer redundant.

6.2.4 Microstructural Studies with a Scanning Tunneling Microscope (STM) and an Atomic Force Microscope (AFM)

STM and AFM studies [4, 5] are possible down to the atomic level. Individual surface atoms can be imaged. STM and AFM techniques have had a major impact on the microstructural studies of surfaces with high resolution. STM can only be

performed on conductive materials, such as TiO_2 and doped ceramics. AFM studies can, however, be applied to nonconductive specimens. Recent instrumental developments have made AFM work very easy and convenient. In particular, grain size and grain size distribution can readily be investigated [14]. Different phases can also be identified, as well as specific problems such as surface structures near interfaces.

6.2.5 Microstructural Studies with a Transmission Electron Microscope (TEM)

Transmission electron microscopy is one of the most important tools for the characterization of property dependent quantities of structural ceramics. There are excellent textbooks on TEM [15, 16], high resolution electron microscopy [17, 18], and analytical electron microscopy [19, 20]. The latter technique enables the determination of the chemical composition with high spatial resolution. Special techniques, such as the study of the energy loss near edge structures (ELNES), allow the identification of bonding states and coordination numbers [20–22].

6.3 Essential Microstructural Aspects of Structural Ceramics

Structural ceramics possess a high melting point leading to extremely high strength, often high corrosion resistance, and low thermal conductivity. The advanced structural materials are mainly used in high temperature engines, but may also be used as refractories [1, 12, 23]. However, the toughness of structural ceramics is rather low; most ceramics are brittle. The critical fracture toughness is much lower than for any type of metal, including cast materials. The deformation diagram shows a steep increase in stress as a function of deformation (strain). At a certain, but usually very high, stress the material fractures catastrophically. This behavior makes it very difficult to use ceramics as structural components, since engineers require the failure of a material to be 'announced' by some form of plasticity. The completely elastic behavior of most ceramics shows that these materials do not deform plastically, but that the failure process starts at weak links in the material. These weak links can be flaws or defects in the material, with emphasis on grain boundaries and interfaces. It is essential to characterize the flaws in monolithic ceramics. Ceramics can be toughened by several processes [24], i.e., by including second phases (composites) or particles ready for martensitic transformation [25]. These materials lead to a quasi-plastic behavior; 'ductilization' is the expression used in the literature.

It should be mentioned that increased reliability of these materials can be achieved by flaw control or toughening. The first approach accepts the brittleness of the material and attempts to control the large processing flaws and all other types of weak defect by microstructural control. The second approach attempts to generate microstructures that impart sufficient fracture resistance that the strength becomes insensitive to the size of the flaw. The former has been the subject of considerable research to identify the most

detrimental processing flaws as well as the processing steps responsible for these flaws. The toughening approach has emerged more recently and has the obvious advantage that appreciable processing and post-processing damage can be tolerated without compromising the structural reliability [26].

The resistance of brittle solids to the propagation of cracks can be strongly influenced by the microstructure and by the use of various reinforcements. An understanding of the properties of both the flaw-controlled and the toughened ceramics actually requires determination of the microstructure using different levels of resolution. For the flaw-controlled materials, the largest flaws have to be identified. This can be done by either scanning electron microscopy or transmission electron microscopy. In these studies the microscopy of the undamaged bulk is usually of no interest and is not relevant. Only the microscopy of defects (and regions close to defects) is relevant for the properties. Therefore these microscopical studies essentially only require the study of flaws and defects.

For the toughened material a similar procedure is required. Only the regions of contact between the different phases (heterophase boundaries) have to be studied. This includes an understanding of the bonding and failure mechanisms at the different boundaries. For some toughened materials plasticity occurs in one phase, e.g., in ceramics toughened by metal particles. For these studies the microscopy of deformed materials is also of great interest.

One final area of importance, more on the experimental side, is quite similar to that involved in the microscopy of metals. In precipitation hardened ceramics or plastically deformed ceramics, the microstructure has to be observed at different levels of magnification. For plasticity studies the concentration of dislocations and other defects is required. For polycrystalline ceramics and composites, the structure and composition control of internal interfaces is crucial. This will be elaborated in the following chapters for specific ceramics and for composites.

6.4 Studies on Alumina (Al_2O_3)

6.4.1 Fundamentals of Alumina

Alumina is the most widely utilized oxide ceramic [27]. It is also the only oxide ceramic widely used in single crystal form. Single crystal alumina, often incorrectly referred to as sapphire rather than corundum, is used for both its structural and optical properties. The only other significant uses of bulk single crystal oxides are artificial gemstones and laser hosts (alumina also finds applications in both of these fields). The overwhelming majority of alumina, however, is used in polycrystalline form. The major markets for alumina-based material on a weight basis are refractories, abrasives, whitewares, spark plugs, and engineering ceramics.

6.4.1.1 Crystallography

There is only one thermodynamically stable phase of aluminum oxide, α-Al_2O_3, which has the corundum structure.

The crystal structure is often described as having O^{2-} anions in an approximately hexagonal close-packed arrangement with Al^{3+} cations occupying two-thirds of the octahedral interstices. The properties of single crystal α-Al_2O_3 are listed in the literature [27–30].

However, many processes, such as the oxidation of metallic aluminum and the thermal decomposition of gibbsite, or vapor-deposited amorphous alumina thin films, involve the formation of intermediate metastable alumina phases. These transition phases are denoted as γ, χ, η, ι, ε, δ, ϑ, and κ and are of particular importance because of their use as catalyst substrates, and because the characteristics of α-Al_2O_3 can be affected by the crystal structure(s) of the precursor(s). All the metastable phases have partially disordered crystal structures based on a close-packed oxygen lattice with varying interstitial Al^{3+} configurations. As equilibrium is approached, the structures become more ordered until stable α-alumina is formed.

The sequence of transition aluminas which form is strongly dependent on the starting material and how it is processed. For example, if the starting material is boehmite, AlO(OH), formed from solution or amorphous alumina, then the sequence is $\gamma \rightarrow \delta \rightarrow \vartheta \rightarrow \alpha$. However, if the starting material is gibbsite, α-$Al(OH)_3$, then the sequence may include $\chi \rightarrow \gamma \rightarrow \kappa \rightarrow \vartheta \rightarrow \alpha$, even though boehmite is formed prior to χ. Diaspore, another polymorph of AlO(OH), transforms directly to α-Al_2O_3. Formation of the transition aluminas from hydrated compounds is often accompanied by the development of a highly porous microstructure [31].

The crystal structures of the transition aluminas are characterized by the maintenance of an approximately f.c.c. anion sublattice [31]. While usually treated as cubic, γ-Al_2O_3 has a slightly tetragonally distorted, defect spinel structure (c/a about 0.99, the distortion varying with heat treatment). δ-Al_2O_3 has a tetragonal superstructure, with one unit cell parameter tripled by the cation vacancies, which are thought to be ordered on the octahedral sites of the spinel structure. ϑ-Al_2O_3 is monoclinic with space group $A2/m$, but it is still similar to spinel and often observed to be twinned; η-Al_2O_3 is cubic spinel.

Alumina has a mixed ionic–covalent bond structure, in comparison with the more covalent silicon nitride and silicon carbide, a lower formation free energy, a high melting point of 2050°C, a specific gravity of 3.99, and its single crystals are transparent.

6.4.1.2 Sintering and Application of α-Alumina

Raw powder is generally produced from bauxite (mainly $Al_2O_3.H_2O$, Al_2O_3-$3H_2O$), which is dissolved in an aqueous NaOH solution in an autoclave at 170°C, then $Al(OH)_3$ is calcined at 1100°C to obtain α-Al_2O_3. γ-Al_2O_3 is formed in a firing temperature range of up to 900°C; for temperatures over 1000°C α-Al_2O_3 forms. The alumina powder so obtained can be pulverized to finer grains and refined to higher purity. Recently, new processes have been developed to prepare very high purity and ultrafine-grained alumina from aluminum alkoxide or aluminum organic compounds, which are

useful for higher grade components and films where higher costs can be tolerated.

Since Al_2O_3 is intrinsically a sinterable material because of its higher ionicity [29, 30], full densification can be achieved by pressureless sintering. However, grain growth occurs simultaneously; this can be inhibited by the addition of a very small amount of MgO dopant. Fine-grained nanocrystalline materials can be obtained via a colloidal route, where interaction of the α-Al_2O_3 particles can be controlled by surface charges. The superior characteristics of Al_2O_3 are its thermal resistivity, electrical insulation, hardness, chemical stability, producibility, and economical cost.

Consequently, an extremely wide range of applications have been developed during the last 50 years. In the high temperature engineering applications there are pipes, plates, and jigs for high temperature uses, transparent tubes for sodium lamps, wear-resistant parts like wire guide and nozzles, mechanical seals, and cutting tools. But this material has a number of major limitations, such as low toughness and poor thermal shock resistance and high temperature strength. Therefore recent development work on this material has aimed to improve its toughness by the addition of other compounds, in the form of particles or whiskers. Hot pressed and hot isostatically pressed (hipped) α-Al_2O_3-TiC have been developed for cutting tools [32, 33]. An addition of TiO_2 is used to inhibit grain growth and to improve the thermal conductivity. α-Al_2O_3-SiC whisker composites have been developed for improved strength and toughness [34, 35].

Alumina strengthened and toughened by ZrO_2 particle dispersions has been developed [36]. The increase in toughness is explained as due to the absorption of elastic strain energy by the formation of microcracks through a martensitic phase transformation from a tetragonal (t) to a monoclinic (m) structure of the ZrO_2. The phase transformation itself can also be utilized for toughening [25]. High strength, high toughness materials have been achieved. Current research is growing in the area of nanocomposites.

6.4.2 Microscopy of α-Alumina

Microscopical studies of α-Al_2O_3 can be subdivided into two main sections. One concerns the determination of the grain size and grain size distribution in fine-grained and ultrafine-grained polycrystalline materials. The other concerns the evaluation of defects with emphasis on grain boundaries. Of course it is also of great interest to study the optical properties of single crystal α-Al_2O_3. However, these studies are not included in this chapter. The reader is referred to the literature [29, 30, 37].

6.4.2.1 Determination of Grain Size and Grain Size Distribution

α-Al_2O_3 possesses anisotropic thermal expansion coefficients which result, after grain growth, in internal stresses upon heating the material. Evans and Fu [38] showed that large grains resulted in stresses that lead to microcracking of the material. The microcracks either toughen the material or they can be the main flaws leading to failure of the material. Therefore it is desirable to maintain a small grain

size in the material. Typically, the grain diameter should be less than 1 μm. Control of the grain size and grain size distribution is essential. Abnormal grain growth [20, 30] has to be avoided.

Since for most alumina materials (including ZrO_2-containing Al_2O_3) the mean grain size is often far below 1 μm, optical microscopy does not yield satisfactory results. SEM and atomic force microscopy (AFM) enable the determination of grain size, grain shape, and grain size distribution [14]. For these studies the surface of the polycrystalline material is polished (see Sec. 6.2.1 of this Chapter) and then thermally etched so that the grain boundaries become visible. The resolution of a high resolution field-emission SEM reaches ~1 nm. Almost atomic resolution can also be achieved with an AFM, because the van der Waals forces utilized in this method vary on an atomic scale. The AFM offers high quantitative resolution in the vertical direction. Moreover, the electrostatic charging effects of insulating materials, as occur with SEM using electrons for imaging, do not occur. An easy and quick investigation of the sample surface is possible in ambient atmosphere. Baretzky et al. [14] recently compared the

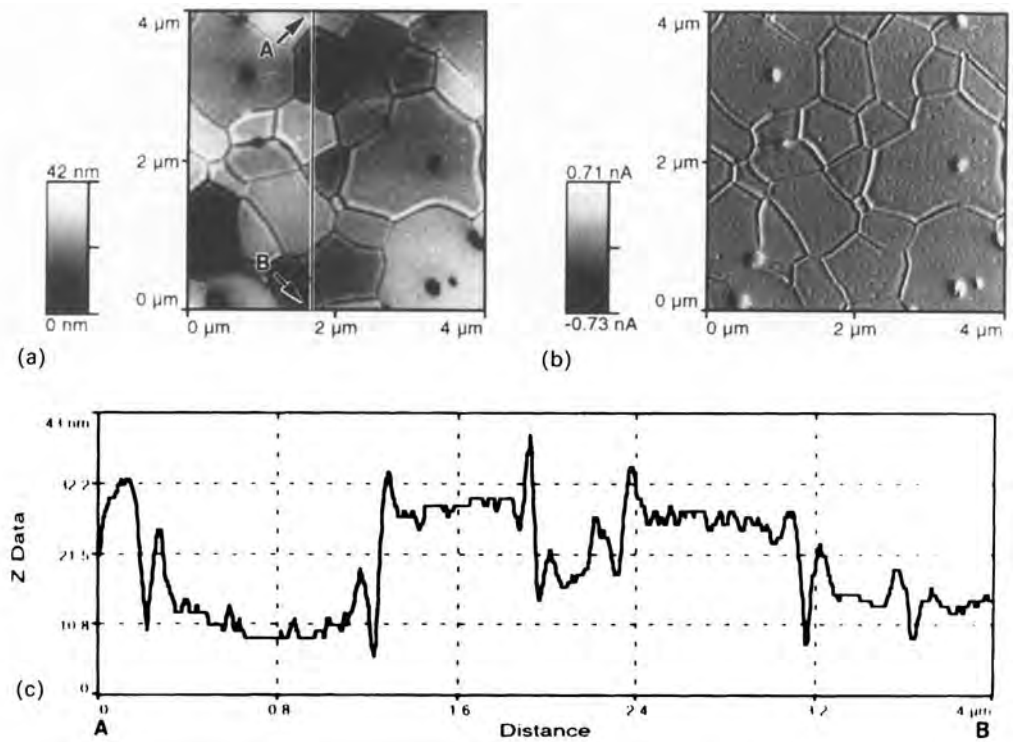


Figure 2. (a–b) AFM micrographs of polycrystalline alumina thermally etched in air at 1200°C for 10 min. (a) Topography image where different heights appear as different brightness values in the color bar. The lowest point was set at 0 nm, the maximum height was determined as 42 nm. (b) Internal sensor image simultaneously recorded, giving the impression of artificial illumination. The measured currents are indicated by different brightness values in the color bar. (c) Quantitative relief profile along line A–B in Fig. 2a. Height data are given as a function of the distance from the origin, defined as point A.

possibility of studying grain size and grain morphology by high resolution SEM (HRSEM) and AFM. The grain size and grain size distribution can easily be determined if the ceramographic preparation is done properly. Figure 2 shows different imaging modes of AFM (topography). With AFM on a very low scale, the atomic microstructure is clearly visible and the size and distribution of grains and pores can be determined. The determination of grain size by optical microscopes is rather difficult.

As an example, in the higher magnification range AFM is compared with high resolution SEM in Fig. 3. The AFM

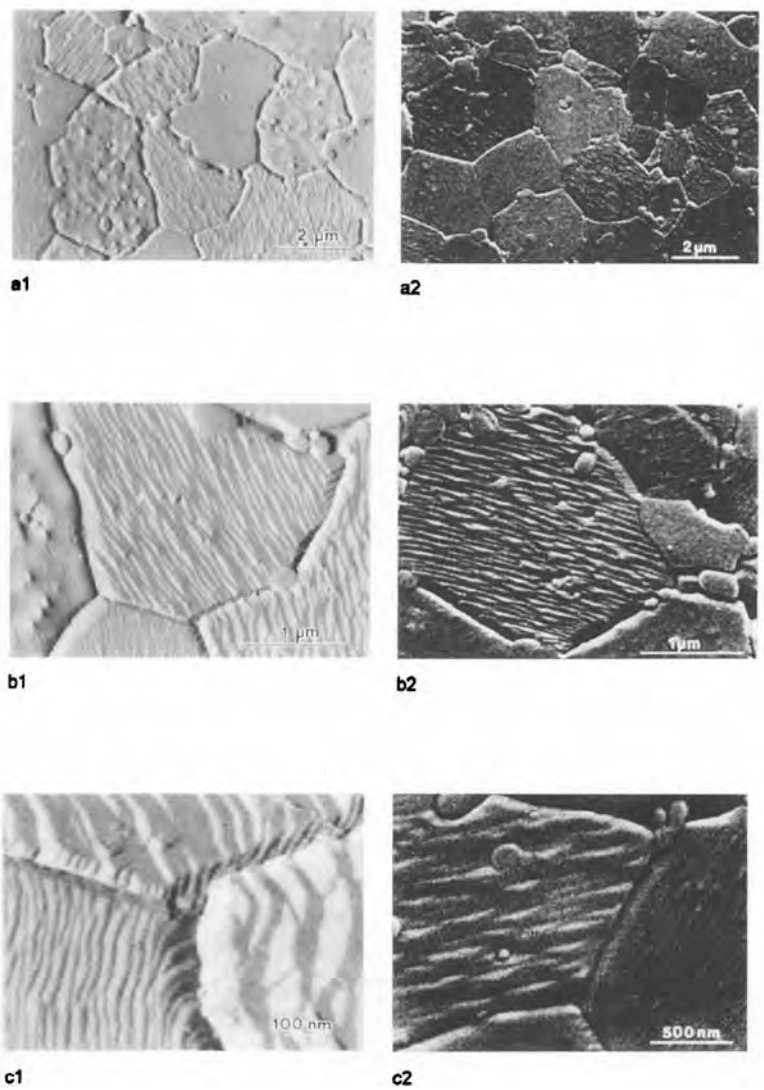


Figure 3. Comparison of AFM (column 1) with HRSEM (column 2) using SiC-doped alumina (sample no. 3) thermally etched under vacuum at 1400°C for 15 min. The magnification is increased from (a) 8000× via (b) 24 000× up to (c) 100 000× for AFM (c1) and 40 000× for HRSEM (c2), respectively.

micrographs are presented in column 1 of Fig. 3 and those of high resolution SEM in column 2. The magnification was increased from ~8000 to 24 000×, up to 100 000× for AFM and 40 000× for HRSEM. The capability of AFM for continuous microscopy, ranging from an overview to the microstructure in one single spot, is demonstrated by the whole series of magnifications, where each micrograph is a magnification of the preceding one. At low magnifications, the microstructural information gained by both AFM and HRSEM is equivalent. Both techniques require ceramographic preparation (thermal etching). Acquisition of the different micrographs usually presents no problem and can be done within a very short period of time. The grain size and grain size distribution can be determined following the standard procedures described, e.g., by Underwood [6].

6.4.2.2 Structure of Undoped Grain Boundaries in α-Alumina

Grain boundaries are often where the failure of a material starts. Therefore it is important to understand the structure and composition of grain boundaries and regions close to the grain boundaries. These studies can only be done on an atomic level by electron microscopy and high resolution electron microscopy (HREM) [39, 40]. Analytical electron microscopy (AEM) yields information on the composition and studies of the near-edge structure in an energy loss (ELNES) result in fingerprints on the bonding and coordination number [21, 22, 41–43].

A bulk of experimental studies of grain boundaries in alumina have been confined to commercial grade polycrystalline alumina, and have been concerned mainly with the crystallographic classification of grain boundaries and the description of grain boundary structures. It is quite interesting that small additions of impurities such as yttrium lead to a change in the distribution of different grain boundary classes [44–46].

Only specific grain boundaries, mainly tilt boundaries, can be studied by HREM [47]. The condition has to be fulfilled that in the direction of the incoming electron beam, strict periodicity occurs within the specimen [48]. (The electron beam has to be parallel to atom columns of the material.) These studies were performed experimentally for different grain boundaries (near $\Sigma 11$ and $1\bar{1}04$ twin) [39, 40]. These grain boundaries were not selected in a technical, polycrystalline, pure material, but from bicrystalline specimens produced by diffusion bonding [49]. The grain boundary misorientation could be adjusted accurately.

The interpretation of HRTEM micrographs requires a comparison between experimental micrographs and simulated images, so that retrieval of the actual structure is possible. This structure can then be compared to results from static lattice calculations. The experimental work by Höche et al. [39, 40] resulted in quite interesting high resolution micrographs (see Fig. 4). A simulated image is included in Fig. 4 as an inset. Reasonably good agreement exists between the experimentally determined structure and the simulated structure. In the interpretation of the micrographs it was possible to differentiate between different metastable structures predicted by theoretical modeling [50]. It was found that the simulated

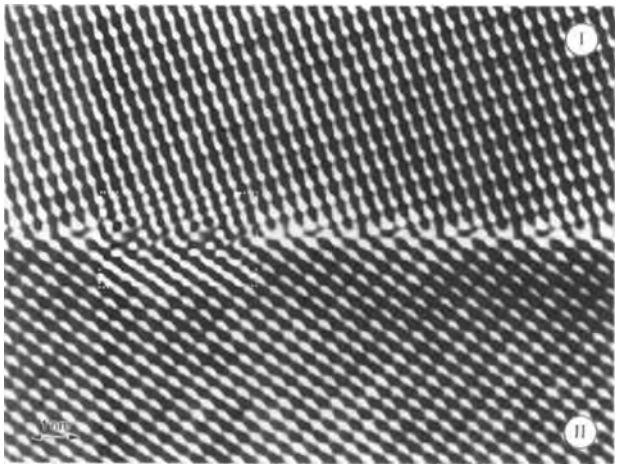


Figure 4. Experimental HRTEM images of the Σ11 grain boundary (Fourier-filtered), [2110] direction; inset: simulated image, thickness ≈ 4 nm, defocus = −5 nm.

images of a specific relaxed structure showed the best agreement with experiments [39]. This agreement was studied quantitatively and it was found that the accuracy of the positions of the columns of ions reached ±0.02 nm. The agreement is not quite satisfactory at the site of the grain boundary. It has not yet been established what causes the difference between the experimental micrograph and the calculated image. This could be caused by deficiencies in the knowledge of the interatomic potentials or by experimental errors (e.g., inhomogeneous TEM specimen, nonequilibrium state of the diffusion bonded grain boundary).

6.4.2.3 Studies of Segregated Grain Boundaries

Alumina has a very low solubility in the bulk for almost all anions and cations [28–30]. Therefore small additions of impurities will easily result in segregation at grain boundaries. Segregated atoms often influence the bonding across the boundaries.

Detailed experimental studies have been carried out on the segregation of calcium doping for a specific grain boundary. These boundaries were produced using ultrahigh-vacuum diffusion bonding of two highly pure single crystals [44]. Two distinct boundary configurations (each having mirror and glide mirror symmetry) were found in the Ca-doped bicrystals. So far only the mirror-related configuration can be identified in the undoped boundary. To estimate the atomistic structure of the grain boundary, simulated images of certain regions of the crystals with different calcium contents were compared to experimental HREM micrographs. The results of these comparisons indicated that the amount of calcium found at this specific twin boundary corresponds to a replacement of ~0.5 monolayers (ML) of aluminum by calcium. These results are comparable to those obtained by analytical techniques.

Recently, great interest has been directed towards studies of the influence of yttrium on the microstructure, segregation behavior, and properties of grain boundaries. It has been established that yttrium

increases the number of special boundaries in bulk α-Al_2O_3 [44, 45]. α-Al_2O_3 scale formed by the oxidation of an aluminum-containing metallic alloy also develops a specific relationship and a columnar grain morphology [51]. Yttrium segregates at grain boundaries and interfaces. Fang et al. [52] showed that yttrium decreases the creep of alumina by about one order of magnitude. The mechanisms of the strong influence of impurities are not yet understood.

The other important impurity for sintering is MgO. MgO reduces the abnormal grain growth of α-Al_2O_3 grain and is therefore essential as a sintering aid. However, different models exist to explain the influence of magnesium. A recently developed model [53, 54] seems to be reasonable for understanding the different mechanisms. However, the atomistic structure of the MgO-containing grain boundaries has not yet been studied by AEM and HRTEM.

6.4.2.4 Grain Boundary Films

Silica-based alumina materials contain an amorphous grain boundary film which wets all boundaries. So far the phenomenon has been studied, along with the wetting of these boundaries [27]. However, detailed consideration of the structure, composition, and behavior of these grain boundary films has not been studied at the same level as in Si_3N_4 (see Sec. 6.6 of this Chapter). These silica-based alumina ceramics are the cheapest materials that can be processed. Their properties are not very good. High temperatures result in a very high creep rate. However, these materials are used as insulators due to their low cost. A detailed analysis of their microstructure reveals that the distribution of the amorphous silica film within the material depends on the exact processing configuration and details of the processing route. Unfortunately, insufficient detailed microscopy studies have been done for this type of material.

Alumina-based ceramics with 95–99% Al_2O_3 have a large proportion of grain boundary glass which must be of carefully controlled composition to confer the required densification behavior and final state properties. The grain boundary glass is usually an aluminosilicate containing additional oxides, such as CaO or MgO. Microstructures of aluminas where the glass is used simply as a densification aid show a uniform distribution of alumina crystals completely separated by glass. In other aluminas, which are fired to higher temperatures, some recrystallization of the alumina may occur to give an interconnected network. In these materials pores are usually located at the interface between the alumina and the glass.

6.5 Studies on Zirconia (ZrO_2)

The traditional applications of ZrO_2 and ZrO_2-containing materials are foundry sands and flours, refractories, ceramic and paint pigments, and abrasives. These applications still account for most of the tonnage used. However, the thermomechanical and electrical properties of zirconia-based ceramics have led to a wide range of advanced and engineering ceramic applications. Early reviews of the then state of knowledge of ZrO_2 are given by Ryshkewitch and Richerson [23] and

Garvie [55]. The recent level of research interest in ZrO_2 can be gauged by examining the series of conference proceedings on the science and technology of zirconia [56–59]. Tough, wear resistant, and refractory, ZrO_2 is being developed for applications such as extrusion dies, machinery wear parts, and piston caps. Composites containing ZrO_2 as a toughening agent are used, e.g., as cutting tools. Ionically-conducting ZrO_2 can be used as a solid electrolyte in oxygen sensors, fuel cells, and furnace elements [27]. TEM played and still plays a major role in revealing the microstructure and the mechanisms responsible for different effects.

6.5.1 Phase Transformations of Zirconia Ceramics

ZrO_2 undergoes several phase transformations upon cooling from the melt [60–62]. At least three crystallographic modifications exist which possess cubic (c), tetragonal (t), and monoclinic (m) symmetry and are stable at high, intermediate, and low temperatures

$$\text{melt} \xrightarrow{2680^\circ\text{C}} \text{cubic} \xrightarrow{2370^\circ\text{C}} \text{tetragonal} \underset{1150^\circ\text{C}}{\overset{950^\circ\text{C}}{\rightleftharpoons}} \text{monoclinic}$$

The high temperature c-phase has the fluorite structure (*Fm3m*), whereas the other polymorphs (t: $P4_2/nmc$, m: $P2_1/c$) are distorted versions of this structure (Fig. 5). The $t \rightarrow m$ transformation is martensitic in nature. The addition of Y_2O_3, CaO, or MgO into the crystal structure maintains the highest temperature configuration (c-ZrO_2) even at low

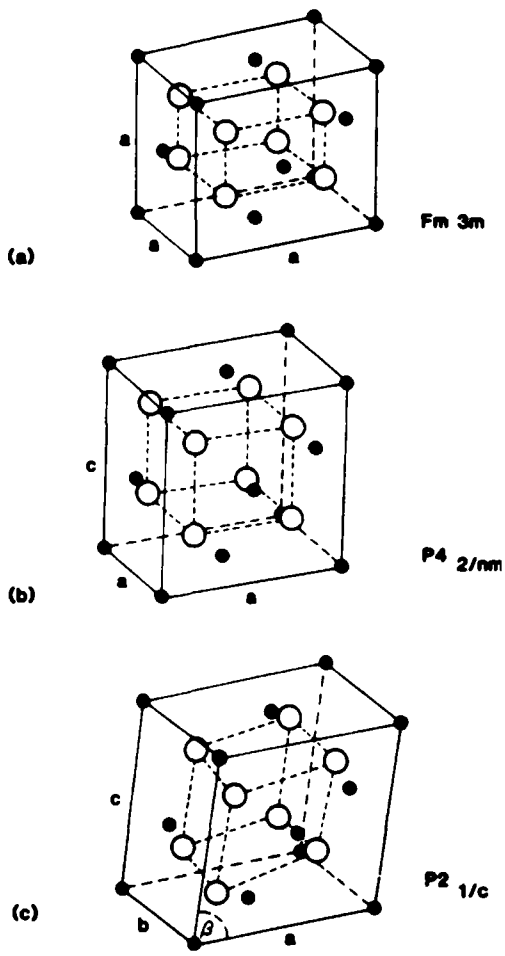


Figure 5. Schematic representation of the three zirconia polymorphs: (a) cubic; (b) tetragonal; (c) monoclinic. Their space groups are indicated.

temperatures. This can be seen from the phase diagram (see, e.g., Fig. 6).

Several types of microstructure can be used as a basis for ZrO_2-toughened ceramics (ZTC). Of particular importance are two different compositions, materials I and II, shown on the phase diagram (Fig. 6). Material I may be densified in the t-ZrO_2 phase field and cooled to room temperature (RT). Then, provided that the grain

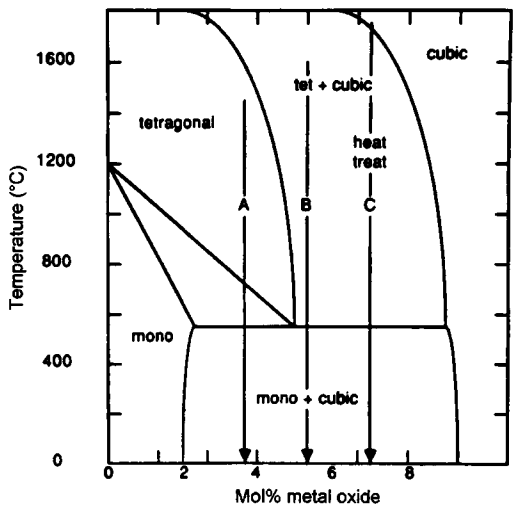


Figure 6. A schematic metal oxide–zirconia phase diagram showing routes for fabricating three different microstructures.

size is less than the critical size required for the transformation, single-phase t-ZrO_2 may be produced (tetragonal zirconia polycrystals: TZP). Composition II (Fig. 6) consists of t-ZrO_2 precipitates in a cubic matrix. Generally, these materials are sintered in the cubic phase field [63–66], but on cooling to the two-phase field, the tetragonal phase forms as precipitates. The size and size distribution of the tetragonal precipitates depends on the thermal treatment in the two-phase field. The t-ZrO_2 precipitates, when small, remain tetragonal upon cooling down to very low temperatures (even liquid helium). This class of materials is referred to as partially stabilized zirconia (PSZ). The shape of the t-ZrO_2 particles depends on the stabilizing agent [56–59]. Conversely, large precipitates transform thermally to m-ZrO_2. Finally, TZP may be used as one of the components in a composite material. Such materials are referred to as ZrO_2-dispersed ceramics (ZDC). The predominant system is Al_2O_3-ZrO_2 (zirconia-toughened alumina, ZTA). The size and distribution of the t-ZrO_2 regions depend on the processing conditions.

The $t \rightarrow m$ phase transformation (Fig. 7) involves a set of transformation strains that increase the volume and change the shape of the particle/grain.

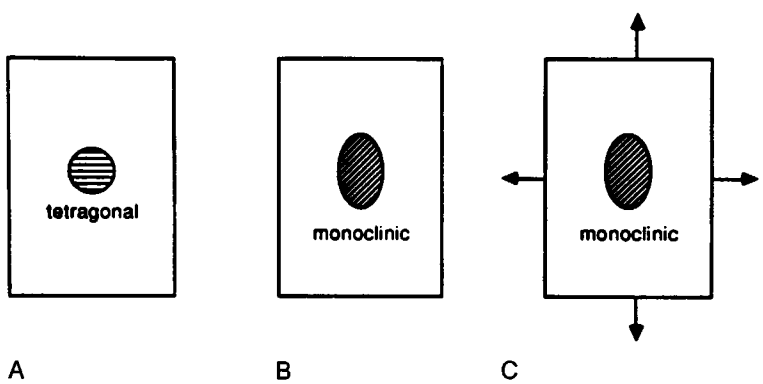


Figure 7. A schematic representation of the zirconia phase transformation. The normal phase transformation to monoclinic is represented by going from A to B; the zirconia particle undergoes a size and shape change as indicated. The material surrounding the particle will oppose the transformation and it is the strain energy that is involved in this constraint that allows the tetragonal phase to be retained. As shown in C, the transformation from A to C can be aided by an applied stress.

Isolated grains invariably transform provided that the nucleation conditions are satisfied. However, for grains embedded in a matrix, strain energy changes tend to oppose the transformation. Then for the transformation to proceed the system has to be 'supercooled' to sufficiently increase the chemical driving force. The retention of t-ZrO_2 to room temperature can be controlled by several microstructural and chemical factors, such as grain size and alloy content.

The microstructure of these materials can be characterized by methods that depend on the level of resolution required for the identification of structures and morphologies. Usually, electron optical instruments are needed. For example, SEM has allowed the determination of the grain sizes and phases in ZrO_2 materials down to the dimensions of about 0.1 μm, whereas TEM has allowed the identification of phases and defects to the atomic level. Some typical TEM micrographs obtained for Mg-PSZ, ZTA (with intercrystalline and intracrystalline t-ZrO_2), and Y-TZP particles are depicted in Figs. 8–10. Major defects in these materials, such as dislocations and precipitates, have also been analyzed by conventional TEM [15], while analytical TEM has enabled the characterization of the chemical composition with a high spatial resolution of ~1 nm [19–21]. Finally, HRTEM has enabled the determination of the atomistic structure, wherein experimentally obtained images have been compared with results of image simulations using specific models [17, 18].

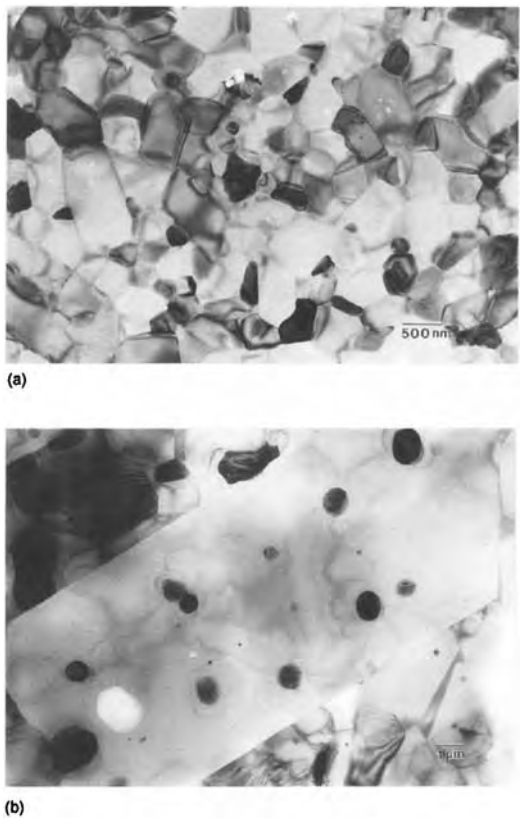


Figure 9. TEM micrograph of an Al_2O_3 ceramic containing dispersed ZrO_2 particles: (a) irregular-shaped, faceted ZrO_2 particles are intercrystalline; (b) Regular-shaped ZrO_2 particles are intracrystalline; the orientation of the particle is random inside the matrix.

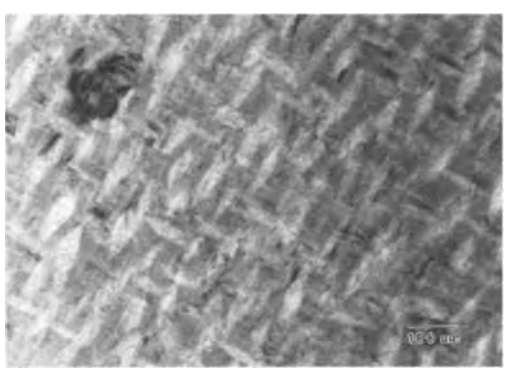


Figure 8. TEM micrograph of an Mg-PSZ ceramic. The grain size of the cubic matrix is ~50 μm. Small precipitates are visible.

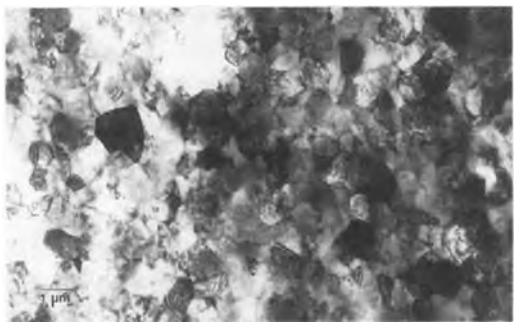


Figure 10. TEM micrograph of a TZP ceramic which is 100% t-ZrO_2. Grain size range <0.8 μm.

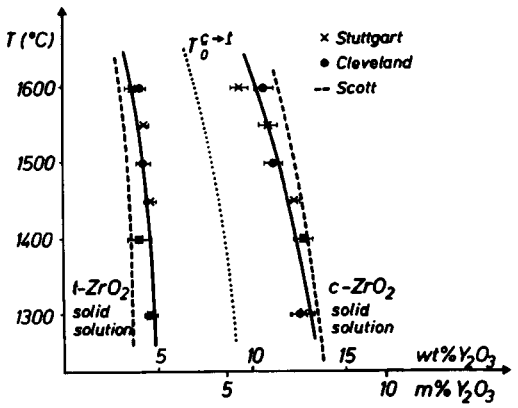


Figure 11. Zirconia-rich end of the yttria–zirconia equilibrium diagram. The equilibrium compositions were determined by energy dispersive X-ray spectroscopy.

6.5.2 Microscopy of the Phase Transformation in Zirconia Ceramics

Many theoretical and experimental studies have been performed with respect to the phase transformations. The nature of the c → t transformation is not yet understood. It is still being debated whether the phase transformation is martensitic or thermally activated.

The details of sections of the phase diagram could be analyzed by AEM. Enhanced precision is illustrated by studies of subsolidus equilibria using AEM to determine the compositions of the coexisting t- and c-ZrO_2 with high spatial resolution and with calibrated standards to determine absolute concentrations [67, 68]. The data (Fig. 11) obtained using crystals containing 6 or 8 wt.% Y_2O_3, grown by skull melting and annealed for long times between 1300°C and 1600°C, have an absolute uncertainty of ±0.2% and represent the best determination to date of subsolidus phase equilibria in a ceramic system in which both coexisting phases show appreciable solid solubility.

The t → m transformation is martensitic in nature [61, 62]. The crystallographic aspects of the transformation are well established. Experimental studies show that the *c*-axis of t-ZrO_2 is parallel to the *c*-axis of m-ZrO_2. The nonvanishing components of the symmetrical unconstrained transformation strain tensor, ϵ_{ij}^T, defined with respect to the tetragonal lattice, for pure ZrO_2 are [69, 70]

$$\epsilon_{11}^T = \frac{a^m}{a^t}\cos\beta - 1 = -0.00149 \quad (1)$$

$$\epsilon_{22}^T = \frac{b^m}{b^t} - 1 = 0.02442 \quad (2)$$

$$\epsilon_{33}^T = \frac{c^m}{c^t} - 1 = 0.02386 \quad (3)$$

$$\epsilon_{13}^T = \epsilon_{31}^T = \tfrac{1}{2}\tan\beta = 0.08188 \quad (4)$$

The shear components have the largest values, while the volume strain for room temperature ZrO_2 is about 4.7%.

Strain fields that occur in association with ZrO_2 particles are important to the transformation, as elaborated below. Consequently, the local strain field in the matrix surrounding a zirconia particle

has been examined by HVEM [70, 71]. Quantitative analysis of the strain field by TEM provided an important contribution to the understanding of the transformation mechanism. The strain analysis was performed at intracrystalline, regular-shaped, spherical, and slightly ellipsoidal t-ZrO_2 particles completely embedded in an Al_2O_3 matrix. Extended contrast fringes are visible around the embedded particles when the matrix is imaged under two-beam dynamical diffraction conditions. By imaging the Al_2O_3 surrounding the particles with different diffraction vectors (Fig. 12) and at different orientations, it was determined that an anisotropic strain field surrounds the t-ZrO_2 inclusions and that the particle is a center of dilatation. The strain anisotropy is attributed to the anisotropic thermal expansion mismatch. The specimens were fabricated in equilibrium at ~1500°C. Preliminary experiments correlating the strain field with the internal orientation of the tetragonal phase are consistent with the hypothesis that maximum misfit with respect to the alumina occurs along c_t [71]. Computer simulated images from contrast calculations based on the above model match well with the observed tetragonal strain field [72, 73]. Additionally, contrast calculations conducted using elastic dipoles yielded misfit parameters which agreed reasonably well with those expected from the misfit strain [70, 73].

Transformed particles are internally twinned, leading to very high displacement fields localized at the particle–matrix interface (Fig. 13). Quantitative analysis indicates that the particle is a center of compression due to volume increase and the fact that the strain field symmetry axis is perpendicular to the internal twin plane.

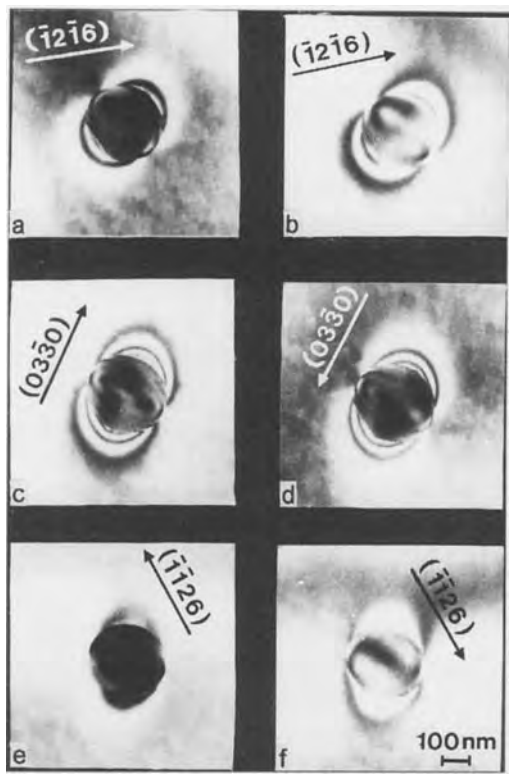


Figure 12. HVEM analysis of the anisotropic strain field in the alumina matrix surrounding an intragranular t-ZrO_2 particle which formed a center of dilatation at room temperature. The particle was imaged in bright dark-field mode under two-beam dynamical conditions with the plane normals (diffraction vectors $\boldsymbol{g}$) indicated.

The transformation and deformation twinning mechanism for a transforming ZrO_2 particle in an alumina, suggested by the observations, is summarized in Fig. 14.

The nucleation of m-ZrO_2 in t-ZrO_2 could also be observed directly by TEM in a fine-grained Y_2O_3-containing t-ZrO_3 polycrystal (Y-TZP). These materials possess a unique microstructure. The t-ZrO_2 grains are $\leqslant 1\,\mu m$ in diameter and constitute $\geqslant 80$ percent of the microstructure; in some samples, c-ZrO_2 grains are present. All grains are surrounded by a thin,

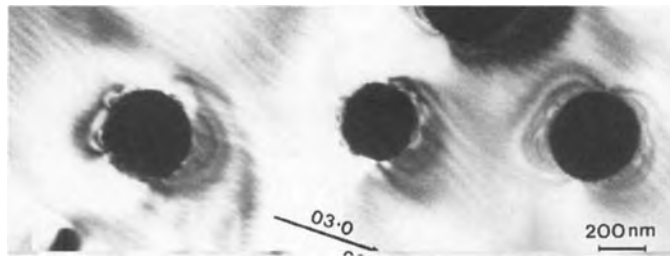


Figure 13. HVEM micrograph of strain fields due to internal monoclinic twins localized at the particle–matrix interface.

continuous amorphous phase [74]. No grain boundary dislocations can therefore be present. Nevertheless, such boundaries are sites of localized residual stresses arising from the thermal expansion anisotropy of t-ZrO_2 [75].

Y_2O_3 is a very effective stabilizing solute and increases the difficulty for martensite nucleation to occur. Among other manifestations of this difficulty, the transformation zone found from in situ crack propagation experiments in the HVEM is discontinuous [76]. The relative stability of t-ZrO_2 in this system facilitates study of the nucleation process. It was shown that t-ZrO_2 can often be made to transform during TEM observations, due to stresses induced by electron beam heating [77]. (ZrO_2 has one of the lowest thermal conductivities among inorganic crystalline solids.) Nucleation is found to occur preferentially at grain boundaries and grain corners, and it is possible to photograph various stages of the transformation. A typical example is shown in Fig. 15. A typical, essentially featureless, high-angle grain boundary is present at the beginning of the experiment. After a short interval of electron irradiation, strain contours develop at a specific site along the grain boundary (the arrowed feature in Fig. 15b). These strain contours oscillate during the experiment and sometimes die away. Occasionally, however, a small lathe of martensite (m-ZrO_2) forms out of these oscillatory contours and begins to grow across the grain (Fig. 15c, d). It is worth noting that the strains close to dislocations, small-angle boundaries, and stacking faults do not act as preferred embryos.

The nucleation in a t-ZrO_2 particle in a precipitation-toughened ceramic is rather complex. The t-ZrO_2 is present as coherent

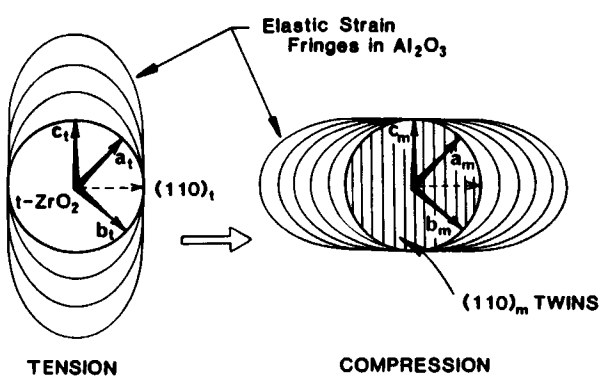


Figure 14. Suggested transformation and deformation twinning mechanism for a transforming ZrO_2 particle in an α-Al_2O_3 matrix.

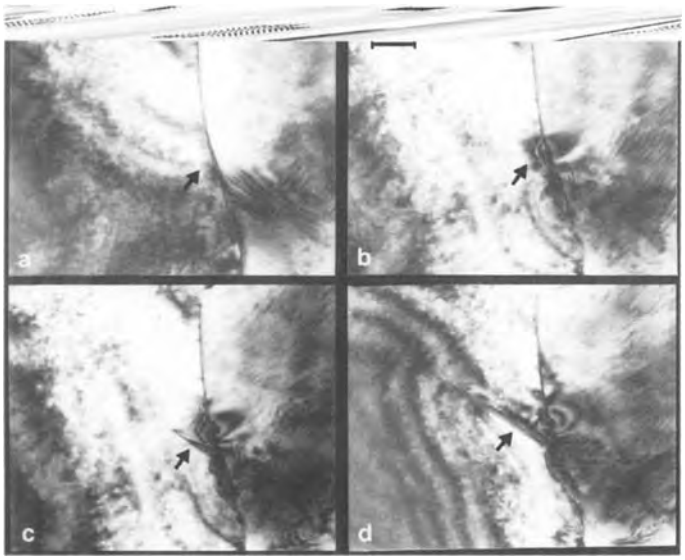


Figure 15. Nucleation in Y-TZP. (a) At the beginning of the experiment, no defects or localized strain contours are visible on the grain boundary; they develop during irradiation, as shown by the arrowed region in (b). On continued irradiation, a martensite lathe (an m-ZrO_2 plate) develops [the arrowed feature in (c)] and grows into the grain (d). Bright field electron micrographs.

precipitates in partially-stabilized ZrO_2 (PSZ), and it has been suggested [78] that the size dependence of M_S is related to the loss of coherency during extended high temperature ageing: the interface dislocations act as sites for heterogeneous nucleation. Indeed, Chen and Chiao [79, 80] concluded that a single screw dislocation with [001] Burgers vector should suffice to nucleate the transformation at room temperature. However, virtually all PSZ ceramics studied to date in the TEM have been essentially dislocation-free; dislocations could not be imaged even at precipitate–matrix interfaces. The role of dislocations as nuclei thus appears ambiguous.

It has been suggested that inhomogeneous internal stresses develop in PSZ ceramics owing to the specific shapes of t-ZrO_2 precipitates [81, 82] in an Mg-PSZ ceramic (Fig. 16): the sharp tips (in comparison to a perfect ellipse) at the end of the particle result in large stresses, which will result in large localized stresses responsible for the nucleation [82, 83].

6.5.3 In Situ Observation of the t → m Transformation in Zirconia

Direct observation of the stress-induced transformation in ZrO_2-containing ceramics was carried out by in situ experiments in a high voltage electron microscope. During deformation, cracks were introduced in a specially prepared TEM

Figure 16. Micrograph of a t-ZrO_2 precipitate in Mg-PSZ with an oblate ellipsoid, having the same principal axes.

Figure 17. Sequence of straining experiments. (a) TEM micrograph taken prior to straining. No m-ZrO_2 inclusions are present. (b) Crack starts to propagate. The crack is marked on all micrographs (c) to (f). Crack propagates on increased loading. t-ZrO_2 inclusion transformed to m symmetry.

specimen where two elongated holes with a transparent ridge between them were formed.

The spatial distribution of zirconia particles in ZTA, which transformed ahead of the crack as it progressed through the material, is summarized in Figs. 17 and 18. The general shape of the transformation zone could be outlined (Fig. 18), although its dimensions were dependent on the microstructure's homogeneity. Clusters of zirconia particles, which were susceptible to transformation even at larger distances from the crack, modified the shape of the zone. The expected transformation ahead of a crack tip within a well-defined transformation zone was thus experimentally verified. Furthermore, the zone height could be evaluated by investigation of the t → m transformation in the 100-μm long wake of the crack [76]. These TEM studies represent direct proof of the stress-induced transformation in t-ZrO_2. The TEM results were used to model the transformation toughening of zirconia-containing composites.

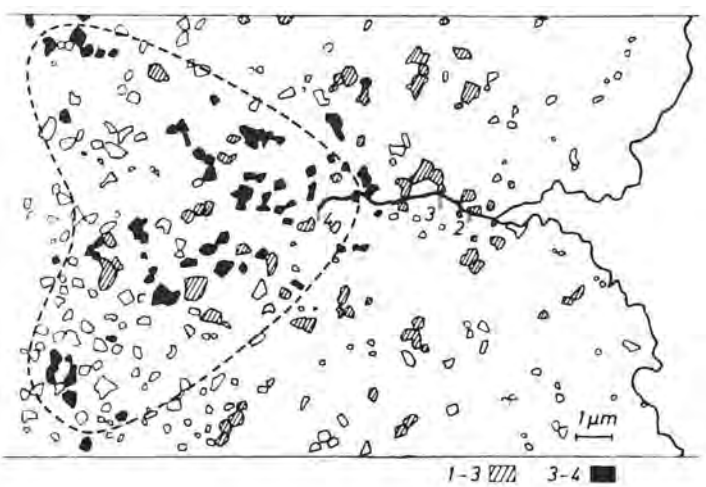


Figure 18. Schematic representation of (partial) evaluation of the straining experiment. The specimen was strained so that the crack propagated to different positions (1, 2, 3, 4). All ZrO_2 inclusions are marked on the schematic drawing. Hatched areas: ZrO_2 grains transformed during propagation of the crack from the hole to position 3. Dark areas: ZrO_2 grains transformed during crack propagation from position 3 to 4. The transformation zone is outlined.

6.5.4 Microscopical Studies of Microcrack Toughening

Buresch [84], Hoagland et al. [85], and Evans [86] postulated the phenomenon of microcrack toughening more than two decades ago and indeed, a range of materials do exhibit trends in toughness in relation to particle size, temperature, etc., qualitatively consistent with this mechanism. However, as yet there is only one fully validated example of this mechanism: Al_2O_3 toughened with monoclinic ZrO_2 [87, 88]. Microcracks occur within regions of local residual tension, caused by thermal expansion mismatch and/or by transformation [89, 90]. The microcracks locally relieve the residual tension and thus cause a dilatation governed by the volume displaced by the microcrack. Furthermore, the microcracks reduce the elastic modulus within the microcrack process zone. Direct evidence of the toughening mechanisms has so far been provided exclusively by detailed TEM studies. Microstructural and microcracking studies were performed on various ZTA materials wherein the proportion of ZrO_2 thermally transformed to the monoclinic phase was varied between 14 and 77%. TEM studies revealed that in these materials all ZrO_2 particles and Al_2O_3 grains are faceted, that the particles are intercrystalline, and that the grain boundaries and interface

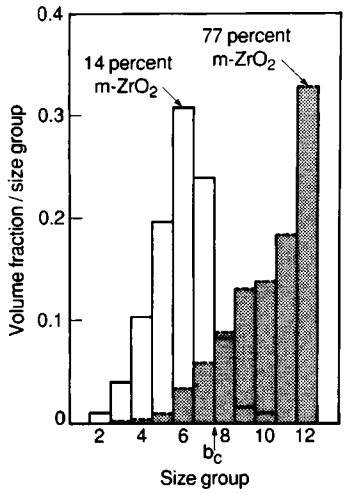


Figure 19. ZrO_2 particle size distribution for materials with 14% and 77% of the ZrO_2 in the monoclinic form. All particles larger than b_c are monoclinic.

boundaries were circumvented by an amorphous layer [91]. The size distributions of the grains and particles have been measured in foils up to 0.8 μm in thickness, with the radii separated into different nonlinear size groups using the volume fraction per size group as the ordinate (Fig. 19). ZrO_2 particles with radii $b_c > 0.3$ μm were found to be monoclinic and contained twin lamellae.

Small interfacial microcracks were usually observed between monoclinic ZrO_2 particles and the matrix grains in those regions where the twin planes of m-ZrO_2 terminated (Fig. 20). Occasional, thermally-induced microcracks were also

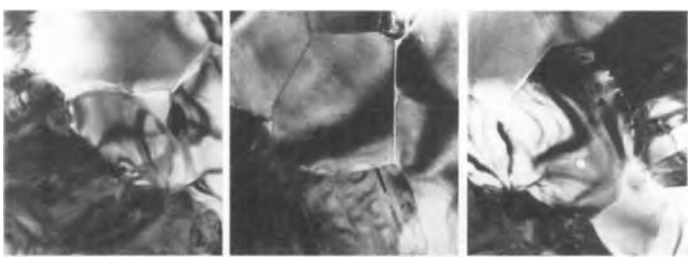


Figure 20. Three examples of microcracks observed at twin terminations in the Al_2O_3/ZrO_2 interface.

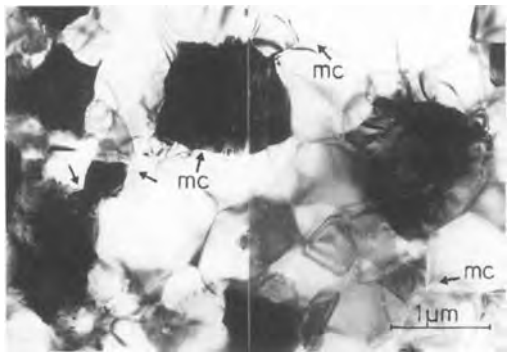


Figure 21. A TEM micrograph revealing typical microcracks (mc) in the process zone.

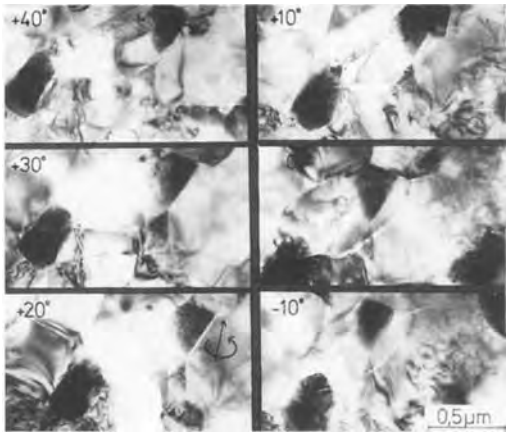


Figure 22. A tilting sequence indicating the technique used to trace microcracks.

detected remote from the macrocrack [87]. These were invariably associated with larger monoclinic particles: the largest having a diameter of ~7 μm. These cracks dominate the strength, resulting in strengths less than those apparent in the predominantly tetragonal material.

Matrix microcracking caused by the growth of macrocracks was studied by preparing thin foils at four distances, y, from a macrocrack (0.5, 1.5, 3, and 6 μm). Radial matrix microcracks were observed, as exemplified in Fig. 21. All such radial microcracks occurred along grain boundaries in the Al_2O_3 matrix. Usually, the interface between the Al_2O_3 and the ZrO_2 was debonded at the origin of the microcrack. The detectability of radial microcracks depended on their inclination with respect to the incoming electron beam. Trends in the visibility of microcracks upon tilting around one axis (Fig. 22) indicated that tilting in all directions would be needed to detect each microcrack present in the foil. However, tilting in the TEM is limited to ±45° in all directions, so that only 0.3 of the solid angle is covered. Therefore the fraction of detectable microcracks is limited to 0.3 ± 0.05.

Subject to this detectability limitation, large regions of TEM foils of known thickness have been investigated. One example is shown in Fig. 23, wherein all

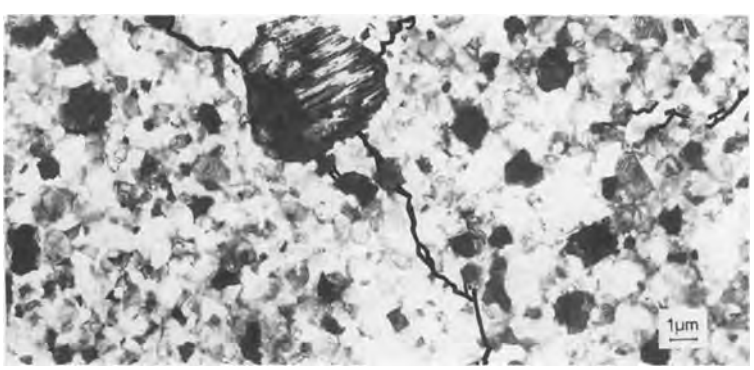


Figure 23. A TEM micrograph with all observable microcracks marked.

microcracks observable under different tilting conditions are marked. The associated projected length, l, of each microcrack has been measured and related to the radius, b, of the originating monoclinic ZrO_2 particle. Additionally, the residual opening of the microcracks, δ, has been determined at the ZrO_2 particle intersection where the opening is usually largest.

The microcrack density could be determined from the micrographs and evaluated using standard techniques of quantitative image analysis [6]. The results of the experiments could be used for an explanation of the toughness increase caused by microcracking [88].

6.5.5 Microscopy of Ferroelastic t′-Zirconia

ZrO_2 reveals, after specific thermal treatment, a ferroelastic behavior [92]. Ferroelastic behavior requires three conditions: (i) A diffusionless phase transition from a high symmetry modification must occur, which generates the ferroelastic material. (ii) The phase transformation must lower the point-group symmetry and must generate a lattice distortion. Usually, the ferroelastic material is composed of at least two different variants (domains) with different orientations. The domains are separated by domain walls. (iii) It must be possible to reorientate the domains by external stresses. The latter property is equivalent to 'switching' the domains by the migration of domain walls.

The t′-ZrO_2 exhibits ferroelastic behavior [93]. The atomistic structure of a domain boundary has been determined [94, 95]. Ferroelastic behavior was produced by compressing quasi single-crystals of t′-ZrO_2 parallel to a [100] direction and the mechanical properties of the material were measured for different deformation temperatures and strain rates [94, 95]. Microstructural evidence of domain reorientation (switching) was observed. Dark-field TEM imaging of the three [112] tetragonal twin variants in a [111] zone was performed before (Fig. 24) and after (Fig. 25) uniaxial deformation parallel to a $[001]_c$ axis. Two sets of twin variants grow at the expense of the third set upon application of uniaxial compression. The diminishing variant was the one with its c-axis parallel to the compression axis. Indentation experiments provided further evidence of domain reorientation near the crack surface. Again, two sets of twin variants grow at the expense of the third set.

6.5.6 Plastic Deformation of Zirconia Ceramics

At elevated temperatures, ZrO_2 ceramics can be deformed plastically [96–98]. The dislocations and dislocation arrangements in the deformed ceramics can be determined by TEM. Fully stabilized c-ZrO_2 usually contains small precipitates and in PSZ ceramics coherent and incoherent precipitates can be identified. TEM observations were performed on plastically deformed specimens. The TEM studies revealed the dislocation arrangements in the materials. Baufeld et al. [99] showed for deformed c-ZrO_2 single crystals containing 11 mol% Y_2O_3 that slip occurs predominantly on the primary slip system,

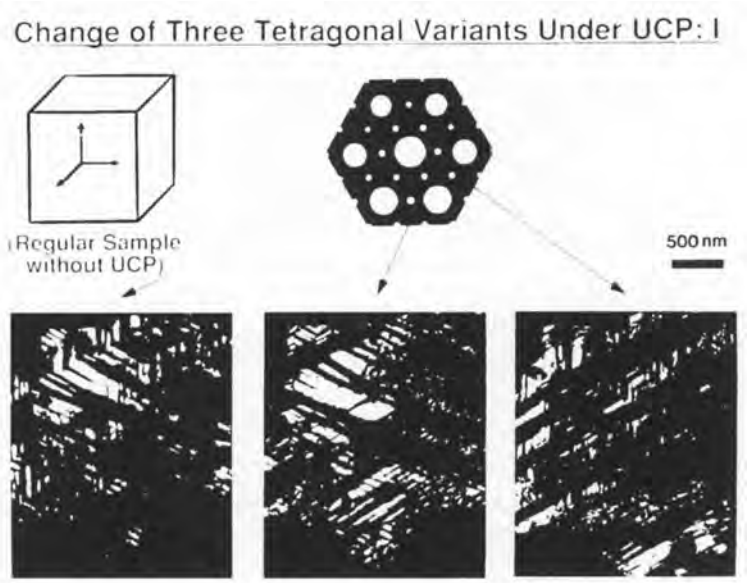


Figure 24. Transmission electron micrographs. Dark-field images of three {112} tetragonal twin variants in a {111} zone axis of an as-heat-treated t'-ZrO_2 material. All three variants are present in the material with about the same probability of 1/3.

and screw dislocations of zig-zag shape within narrow slip bands dominate the microstructure. The strain rate and temperature sensitivities of the flow stress at temperatures ~700°C can be interpreted by the interaction between dislocations and the small (tetragonal) precipitates [99]. Deformation at 1150°C, however, results in serrated flow and the deformation possesses a thermal character [99].

In Y-PSZ ceramics the mechanical properties are determined by the size,

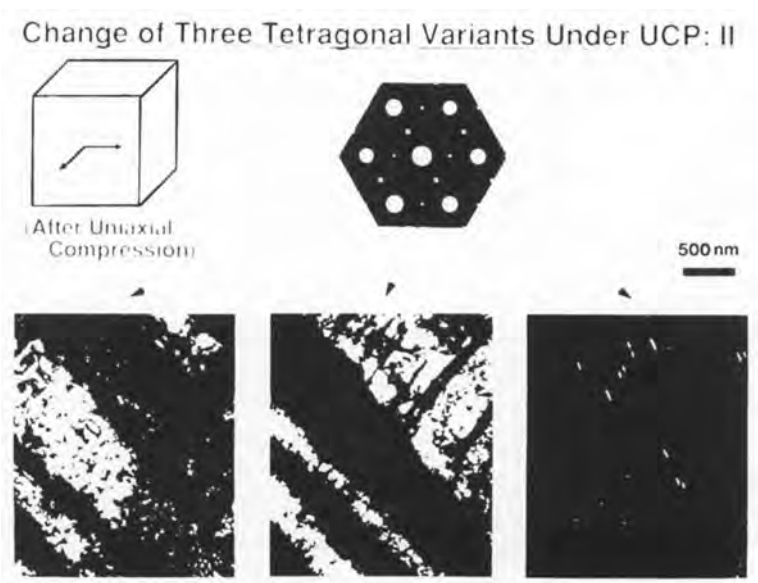


Figure 25. Dark-field TEM images of three {112} tetragonal twin variants in a {111} zone axis of t-'ZrO_2 after the application of uniaxial compression at 600°C. The variant with the *c*-axis parallel to {100} has vanished.

density, and morphology of the precipitates. The morphology of the precipitates can be analyzed by TEM [100]. Cheong et al. [98] pointed out that 'cutting' of the precipitates by a dislocation is anisotropic owing to the lower symmetry and smaller numbers of glide systems. A detailed explanation of the mechanical properties out of a particular microstructure is not yet possible.

6.6 Studies on Silicon Nitride (Si_3N_4)

Silicon nitride is a typical covalent compound with an ionicity of 0.3. It has a higher formation free energy than either alumina or silica, it also has no melting point under ordinary atmospheric pressure, a high vapor pressure, and a very low diffusion coefficient [101, 102]. These intrinsic properties of silicon nitride suggest that this material has poor sinterability to full density. Therefore special sintering technologies such as reaction sintering, hot pressing, gas pressure sintering, and hipping have been developed for this material [103–105]. It is also notable that silicon nitride is easily oxidized in an oxygen-containing atmosphere. As a result, the surface of silicon nitride is always covered by a thin silica layer.

The densification of Si_3N_4 is described as a liquid-phase sintering process. At higher temperatures, SiO_2, which is always present at the surface of the Si_3N_4 particles [106], reacts with the oxide additives to form an oxide melt and, with increasing temperature, an oxynitride melt by dissolving Si_3N_4. While the α-Si_3N_4 particles dissolve and supersaturate the liquid phase, β-Si_3N_4 is reprecipitated [107]. Depending on the composition of the sintering aids, the liquid phase can form an amorphous or a crystalline grain boundary phase during cooling, both of which degrade the mechanical properties of the Si_3N_4 at temperatures >1000°C, because of softening of the grain boundary regions.

For this material, it is essential that its microstructure be controlled by different microscopy techniques. Its microstructural development has to be followed by adequate techniques. These encompass all aspects of microstructural developments, such as nucleation and grain growth.

6.6.1 Nucleation and Growth

Lange [108] investigated hot pressed Si_3N_4 with 5 wt.% MgO and found that the aspect ratio of the grains, and therefore the fracture toughness, was determined by the ratio of α- and β-Si_3N_4 in the starting powder. The strong influence of the α/β ratio on pressureless and gas pressure sintered samples has also been reported by several other research groups [105, 109].

Figure 26 shows the relationship between the number of β-particles in the initial powder (β-nuclei density) and the particle density after complete densification and transformation into the β-phase for five commercial Si_3N_4 powders densified by pressureless sintering with the same amount of Y_2O_3/Al_2O_3 additives [110]. The β-nuclei density was calculated from the β-content determined by X-ray diffraction and the measured particle size distribution of the starting powder under the assumption of the same particle size

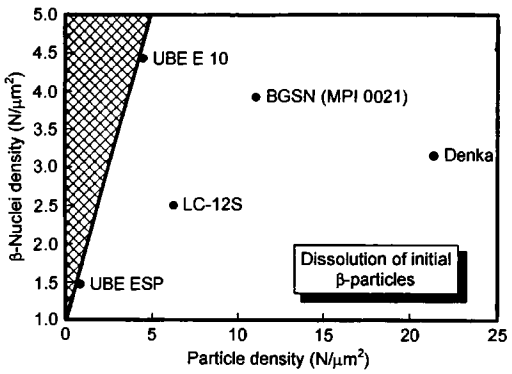


Figure 26. Particle density after pressureless sintering as a function of the calculated β-nuclei density for five Si_3N_4 powders.

distribution for α- and β-particles. The particle density was determined by quantitative microstructural analysis without taking into account the 3-dimensional random distribution of prismatic Si_3N_4 grains. Therefore the measured particle density is somewhat lower than the actual particle density [110].

The hatched area on the left side of the diagram represents powders with a particle density higher than the initial β-nuclei density. This case would imply an additional nucleation either homogeneously or heterogeneously on the α-particles. Powders located on the 1:1 line have the same β-nuclei and particle densities. In other words, all β-particles present in the starting powder grow by dissolution of α-Si_3N_4 particles and reprecipitation as β.

The diagram shows that all investigated powders had an equal or lower particle density in comparison to the initial β-nuclei density. While the UBE powders had approximately the same particle and β-nuclei densities, the particle densities of the BGSN (Bayer AG, Germany) and LC12-S (H.C. Starck, Germany) are smaller than the β-nuclei density, indicating that smaller β-nuclei dissolve during densification. The Denka powder is a β-rich Si_3N_4 powder (97.5%) and complete densification of this powder only occurs by dissolution of the smaller β-particles instead of dissolution of α-grains and reprecipitation on pre-existing β-grains.

A low amount of initial β-nuclei results in a larger interparticle distance of β-grains in the green compact; the grains were able to grow in the first stage of densification without steric hindrance. Therefore the powder with the lowest β-nuclei density (UBE ESP) developed the coarsest microstructure (lowest particle density) with large elongated grains. With increasing initial β-nuclei density, the steric hindrance of grain growth increases, resulting in an increasingly equiaxed microstructure (Denka). The specimens had to be plasma etched prior to the SEM investigation.

By stereology techniques the grain length and aspect ratio distributions were determined for various Si_3N_4 powders. The particles that contain grains with large aspect ratios were intrinsically toughened. Some grains with a large initial diameter can grow in the length direction with minor steric hindrance, more minor than the smaller ones, as pointed out earlier. Their maximum length could reach 100–200 μm. The amount of such large grains depends on the initial β-particle size distribution of the starting powder [110, 111].

6.6.2 The Grain Boundary Film Thickness

TEM investigations are essential for determining the geometrical and chemical

characteristics of the grain boundary film. Clarke [91] was the first to point out that practically all grains were covered with a continuous film consisting primarily of SiO_2 and the sintering additives. The composition of the film is regulated by the thermodynamical phase diagrams [108, 110]. Small angle grain boundaries as well as 'special' grain boundaries are not covered with the film owing to their low intrinsic energy [112]. The typical film thickness ranges from ~0.5 nm to 2.0 nm. There are several imaging techniques however; HREM investigations result in the best measurable film thicknesses [113].

Kleebe et al. [113], Tanaka et al. [114], and recently Pan et al. [115] studied the film thickness with high reliability in different materials. The reliability for the thickness measurements was in the range of ±0.1 nm. The investigations showed that the film thickness was constant for one type of material.

6.6.3 Dependence of the Film Thickness on the Intergranular-Phase Chemistry

The verification of an equilibrium-film thickness implies that each material, with its given secondary phase composition, shows a characteristic film thickness. Previous studies by the authors on Si_3N_4 sintered with different additives in fact revealed a strong dependence of the intergranular-film thickness on the chemistry. For example, the two different materials studied (MgO-doped Si_3N_4 and Yb_2O_3-doped Si_3N_4) were densified with different additives and therefore have a different grain-boundary film chemistry. A consequence of this great difference in the amorphous-film chemistry is the difference in film thickness observed by HREM for these materials [116]. Minor amounts of impurities are expected to have an effect on the intergranular film thickness as well.

The effect of a small amount of impurity on the film thickness was studied by preparing two Yb_2O_3-doped Si_3N_4 materials containing a small amount of CaO (0.25 vol.%) as the controlled impurity [117]. In one material the secondary phase was completely amorphous, while in the other a post-sintering heat treatment resulted in crystallized secondary phases at multiple-grain junctions. Two crystalline secondary phases were observed, $Yb_2Si_2O_7$ and $CaYb_9(SiO_4)_6ON$ (Ca- and N-stabilized apatite). Both materials, sintered and annealed, showed wider grain-boundary films than the undoped material. The annealed material showed a slightly wider film than the as-sintered material (1.4 nm vs. 1.3 nm), which is attributed to a higher calcium concentration at the grain boundaries, as crystallization at triple junctions reduced the overall glass volume in the sintered body, and hence calcium segregates to the boundary. Obviously, the formation of a Ca- and N-stabilized apatite did not completely compensate for the chemical shift in the specimen during annealing. Therefore a slightly higher calcium concentration is expected in the amorphous grain-boundary films in the annealed material than in those in the CaO-doped Si_3N_4 that did not undergo the heat treatment. A similar dependence of film thickness on impurity concentration was also observed in two materials undergoing oxidation [116]. During oxidation the intergranular

phase is essentially purified as additive and impurity cations diffuse to the surface oxide layer [118], resulting in a compositional change that is reflected by a decrease in the grain-boundary film thickness.

6.6.4 A Model System: Low Calcium-Doped Amorphous Silica (SiO_2) Films in Silicon Nitride

Small amounts of CaO additives result in a significant increase in the equilibrium thickness of the intergranular film. This thickness variation may reflect the modification of the atomic structure of the film by the change in chemical composition of the glass. A model system consisting of Si_3N_4 with a pure SiO_2 intergranular phase that was doped with low amounts of calcium was chosen to obtain a better understanding of the influence of impurities of grain-boundary films. Dedicated scanning transmission electron microscopy (STEM) is required for the determination of elemental compositions with a high degree of spatial and energy resolution. The 0.4-nm probe size of the VG HB501 STEM was necessary for elemental analysis of intergranular films of the order of 1 nm in thickness in these materials.

High purity α-Si_3N_4 powder (SN-E10, Ube Corp., Tokyo, Japan) can be sintered

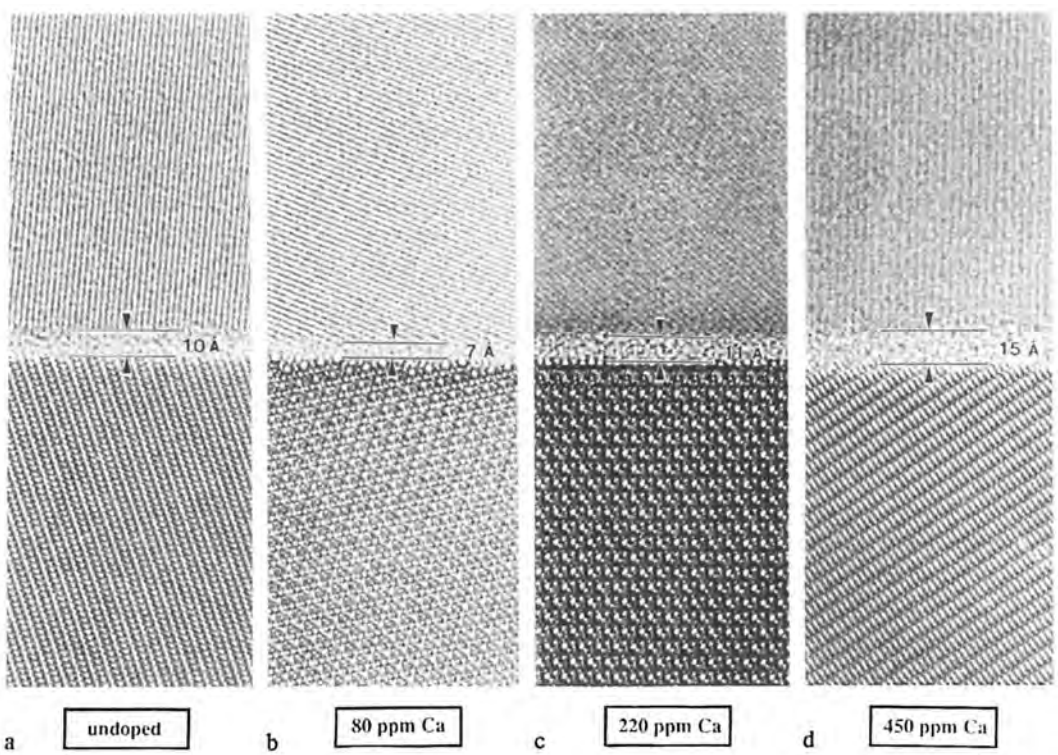


Figure 27. HREM images of grain boundaries in a high purity Si_3N_4 (a) containing an SiO_2 intergranular phase and (b) doped with 80 ppm Ca, (c) 220 ppm Ca, and (d) 450 ppm Ca. Equilibrium film thicknesses were 1 ± 0.1 nm, 0.7 ± 0.1 nm, 1.1 ± 0.1 nm, and 1.5 ± 0.1 nm.

without additives by hot isostatic pressing. The powder contains ~1.3 wt.% oxygen in the surface oxide (SiO_2), but less than 50 ppm of total cation impurities, resulting in a densified Si_3N_4 ceramic with an SiO_2 amorphous grain-boundary phase. Four materials were prepared in this manner, but with controlled amount of calcium (450 ppm, 220 ppm, 80 ppm, and undoped) added to the α-Si_3N_4 powder.

HREM images of these materials with low levels of CaO dopant are shown in Fig. 27. Grain-boundary films in these materials also revealed an equilibrium thickness. The grain-boundary film thickness in Si_3N_4 containing pure SiO_2 amorphous phase was also examined in another undoped Si_3N_4 (prepared from a different Si_3N_4 powder) and in an Si_3N_4-SiO_2 material with a higher SiO_2 content; the film thickness was found to be 1.0 ± 0.1 nm in all of the materials. The presence of calcium was detected in the grain-boundary films of all the materials doped with calcium [119]. Assuming that all the calcium and oxygen are confined in the intergranular phase, the composition of the film can be estimated from the intensity ratio of the calcium and oxygen peaks using electron energy loss spectroscopy. In the 80 ppm Ca-doped material, 1.5 ± 0.3 mol% of CaO was found in the amorphous SiO_2 phase. This value shows good agreement with the simple estimate for the glass containing 1.0 mol% CaO, assuming that all the calcium atoms are localized at the intergranular glass, which is 3.6 vol.% of the material. No impurities were detected in the film in the undoped material, indicating the CaO was present at less than the detection limit of 0.08 mol% CaO. Film thickness as a function of calcium content is plotted in Fig. 28.

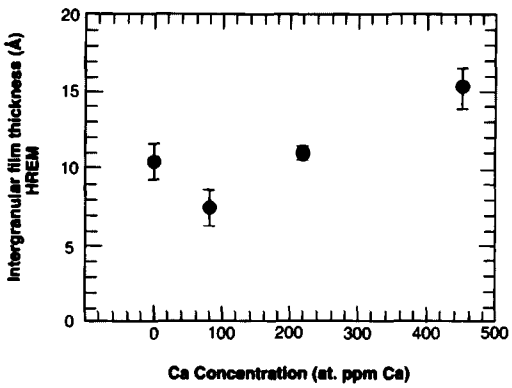


Figure 28. Plot of equilibrium film thickness in high purity Si_3N_4 as a function of the calcium content. The experimentally determined CaO concentration in the intergranular film was determined by high resolution, electron energy loss spectroscopy.

It is interesting to note that a decreasing film thickness in the very dilute region (to a minimum value of 0.7 nm at 80 ppm Ca) is followed by an increase in the thickness as a function of the calcium content.

6.6.5 Interpretation of the Experimental Results

The observed variation in thickness of the intergranular film with the concentration of CaO added can be understood in terms of the forces discussed by Clarke [112] and Clarke et al. [120], namely the attractive van der Waals force and the repulsive steric and electrical double layer forces. In the absence of CaO, we can assume that there is no electrical double layer at the grain boundaries, and so the equilibrium thickness is controlled by a balance between the van der Waals and the steric forces; the resulting grain boundary structure is shown schematically in Fig. 29b to consist of a pure silica glass intergranular

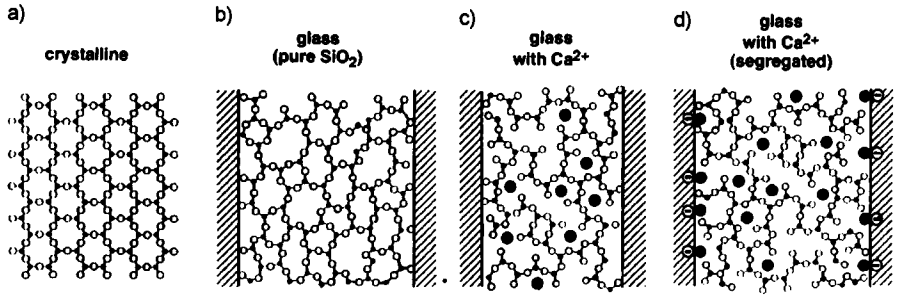


Figure 29. (a–d) Schematic illustration of the changes occurring in the glass structure of the intergranular phase with increasing cation impurity (Ca^{2+}: •) concentration, and its influence on the equilibrium film thickness. A combination of steric and electrical double-layer repulsive forces is proposed to balance the attractive van der Waals dispersion force.

film. The addition of calcia has a dual effect. It is expected to disrupt the network structure of the silica intergranular phase, decreasing the structural correlation length and hence the magnitude of the steric repulsive force, as shown in Fig. 29c. The addition of CaO also provides a charged species to the material which may result in the formation of an electrical double layer, as shown in Fig. 29d. Since CaO is known to be a potent silica network modifier, causing a decrease in the viscosity of several orders of magnitude for additions of less than 1 mol%, it might be expected that the strongest effect of minor calcia additions will be to decrease the steric force and cause a reduction in the film thickness. Further additions of CaO can be expected to increase the repulsive electrical double layer force, leading to further increases in the film thickness. A minimum in the thickness of the intergranular film at a low concentration of a network modifier, as observed at ~80 ppm Ca, is thus consistent with competition between the effects of disrupting the network structure and the development of a force due to an electrical double layer. While more detailed calculations need to be performed to quantify these competing effects, it is nevertheless interesting to note that if all the CaO in the 450 ppm Ca-doped material were adsorbed to the surface of the Si_3N_4 grains, ~40% of the Si_3N_4 surface would be covered by a 0.2-nm thick monolayer, assuming 1-μm diameter grains. If the charge of each Ca^{2+} ion was uncompensated, this would correspond to a rather high surface charge density.

6.6.6 Concluding Remarks

High resolution electron microscopy has an essential and important influence on the development of different Si_3N_4 ceramics. The properties and failure of the materials are not determined by the bulk material but by the composition and structure of defects such as grain boundary phases and triple junctions.

6.7 Studies on Composites

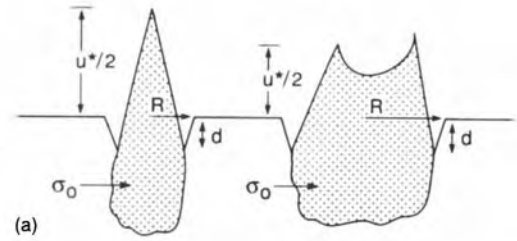
Composites are becoming a more and more important class of materials. They

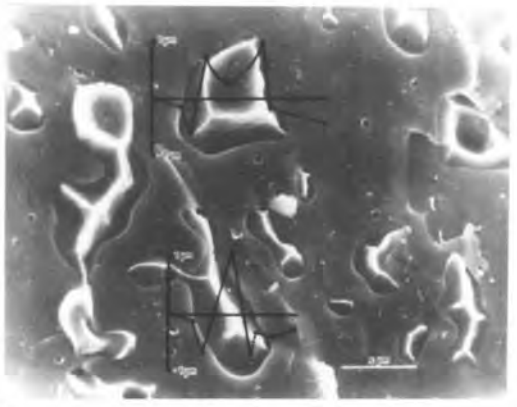
consist of (at least) two components. For a ceramic matrix composite the second (minor) component can be either a metal (metal alloy) or a ceramic.

6.7.1 Al/Al_2O_3 Composites (Lanxide Material)

Of great interest are composites reinforced by metals such as aluminum, copper, or intermetallic alloys. The microstructure depends on the processing route. Those materials processed after the lanxide techniques were investigated by SEM and TEM, and the following salient features were observed. Conventional TEM analyses provide information on two important features: precipitates and dislocations [24, 26]. The size distribution of the precipitates can be determined by TEM; sometimes the precipitates possess a bimodal distribution. The larger precipitates (median diameter ~100 nm) contribute to void nucleation upon ductile fracture of the metallic inclusions. The smaller precipitates (median diameter ~20 nm) are potent dislocation pinning sites. AEM has been used to provide information concerning the character of the precipitates. The larger precipitates contain copper, silicon, and iron in addition to aluminum with trace amounts of chromium. The smaller precipitates consist predominantly of copper and aluminum. In this case no segregation of any impurity is observed at the interface.

SEM investigations of the fracture surface have provided quantitative information about the plastic stretch and the extent of debonding (Fig. 30a). Also various modes of ductile failure have been





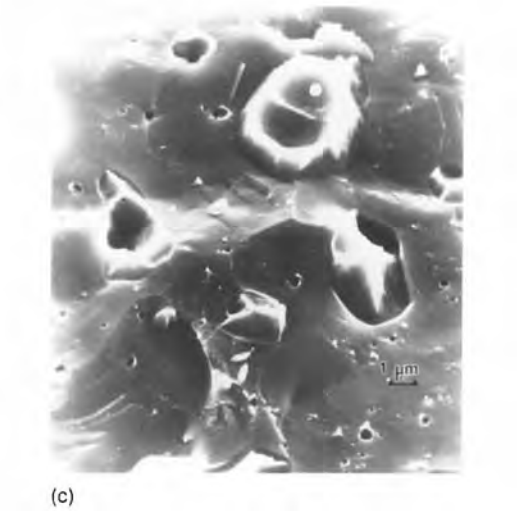


Figure 30. (a) A schematic diagram of the ligament cross section after failure. (b) A scanning electron micrograph of a fracture surface with the topology of the ductile ligaments (as evaluated by stereo microscopy) superposed; this is a typical failure of a ligament having a circular cross section. (c) SEM of a large precipitate at the base of a hole.

identified, as also observed on lead/glass. Specifically, the failure of axisymmetric reinforcement zones often involves the nucleation of a single hole of the center of the neck which rapidly expands to failure (Fig. 30b). In some instances, particles are apparent at the base of the holes (Fig. 30c). These particles are ~100 nm in diameter and probably coincide with the larger Al/Cu/Fe/Si precipitates identified by TEM.

Stereo measurements used to evaluate the plastic stretch to failure, u^*, and the interface debond length, d, revealed that the normalized plastic stretch, u^*/R, varied appreciably between ligaments, ranging from $u^*/R \sim 0.8$ to 2.8. However, there was no systematic dependence on either the ligament dimension R or the aspect ratio. All ligaments have thus been used to provide a cumulative distribution, such that the mean stretch is $\overline{u^*/R} = 1.6$. The debond length also has appreciable variability and ranges between $d/R \sim 0.1$ and ~0.5, but there is no discernible dependence on the aspect ratio. The median value is 0.2. The quantitative experimental data can be used to explain the properties of the material as a result of the microstructure [24].

6.7.2 Al/Al_2O_3 Composites (Processed by Infiltration)

The samples were prepared by slip casting alumina powders [42]. Slurries with 45 vol.% solid were cast into plates, dried, and sintered at 1450°C and 1650°C for 1 h for two different samples. This process resulted in porous preforms with 25 vol.% open porosity. After sintering, the porous plates were infiltrated with pure aluminum (99.999%) at 1050°C with an argon pressure of 15 MPa, maintained for 30 min. For further details of the alumina preparation and melt-infiltration process, see Prielipp et al. [121].

The effects of impurities and dopants on the sintering and microstructure of polycrystalline alumina have been an important issue for some time. Studies have focused on the possible roles of dopants (or impurities) in promoting sintering and limiting grain growth, with specific attention to magnesium [53, 54]. Presumably, it is the segregation of dopants to grain boundaries and surfaces (pores) that influences sintering and grain growth, but the exact mechanism is still not clear. The experimental difficulties in the detection of some segregates further complicates the problem of identifying the roles different dopants play in sintering and grain growth.

Recently, silicon and calcium were confirmed as promoting abnormal grain growth in α-Al_2O_3 [54]. Calcium is a common impurity in alumina, and calcium segregation to general grain boundaries has been experimentally confirmed by both energy-dispersive X-ray spectroscopy (EDS) in a dedicated STEM and Auger analysis. However, recent TEM/EDS studies have shown no evidence of calcium segregation to basal twin boundaries.

Impurity segregation to the free surfaces of alumina is also an important issue, possibly affecting pore mobility, subsequent bonding of surface films to sapphire substrates, or influencing contact angles during liquid-phase wetting. Calcium has been found to segregate to prismatic free surfaces [122, 123]. However, calcium is thought to segregate to

free surfaces anisotropically with no calcium segregation to the basal surface. In the present work, HRTEM and AEM have been used to study (0001) α-Al_2O_3/Al interfaces in melt-infiltrated polycrystalline alumina composites. Liquid aluminum was infiltrated into a porous alumina preform, thus embedding the pores' surfaces and allowing their investigation by TEM. It will be shown that calcium does segregate to the basal surface and to basal twin boundaries, and that it forms a surface phase on external surfaces.

Figure 31 shows a bright-field TEM micrograph of the general specimen morphology. Since all of the neighboring aluminum grains have a similar contrast, Kikuchi electron diffraction was used to check for local texture. All of the aluminum grains located around the 150-μm diameter specimen perforation were found to be in the same orientation. Therefore it can be concluded that after infiltration the aluminum solidifies at a limited number of nucleating sites, resulting in large aluminum grains which extend throughout the open alumina porosity. The alumina shows an elongated morphology with the basal plane parallel to the long axis of the grains. Thus many of the α-Al_2O_3/Al interfaces are parallel to the basal surface and these were studied in closer detail.

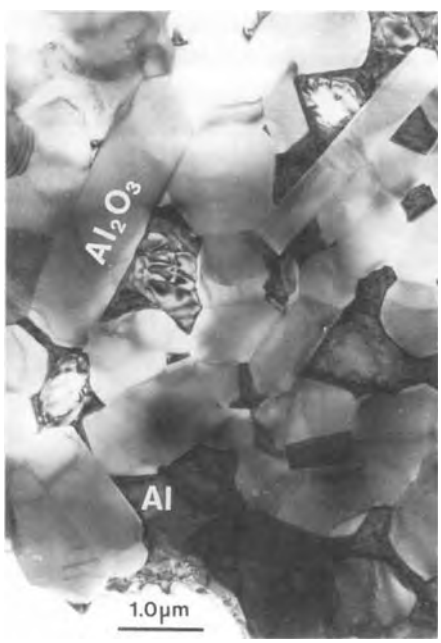


Figure 31. Bright-field micrograph of the microstructure of the melt-infiltrated alumina preform. The aluminum is in the [110] zone axis and appears darker than the randomly aligned alumina matrix grains. Elongated alumina grains are evident.

No wetting defects (open pores) were found between the aluminum and alumina, indicating good wetting of the ceramic by the liquid metal. While this could be explained by the argon pressure used to infiltrate the aluminum into the ceramic matrix, aluminum does not wet the clean (0001) α-Al_2O_3 surface (contact angle greater than 90°), so the lack of wetting defects at the interfaces is surprising.

Figure 32 shows an HRTEM micrograph of a typical basal α-Al_2O_3/Al interface. The alumina grain is in the $[1\bar{2}10]$ zone axis and the interface is parallel to the basal plane, with no steps or facets. The aluminum grain in Fig. 32 is near to a [110] zone axis, but local bending and the presence of dislocations allows only the (002) planes to be visible. The (002) planes are also slightly inclined to the interface plane, and by following them to their termination the transition from bulk aluminum to the interface can be determined. The transition from defect-free α-Al_2O_3 to the interface is more difficult to discern. This can be determined by careful inspection of the (0006), $(\bar{1}012)$, and $(10\bar{1}4)$ lattice fringes and the variation in image contrast from the bulk

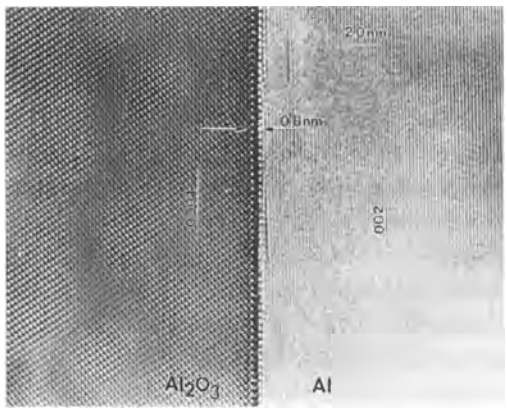


Figure 32. HRTEM image of an (0001) α-Al_2O_3/Al interface. The alumina grain is in the [1210] zone axis, and the aluminum is near to [110]. The width of the interface region is marked by arrows.

regions. Directly at the interface the pattern of the alumina lattice varies significantly from the bulk. This indicates that the symmetry of the relatively wide interface region is different from that of bulk alumina and certainly different from that of aluminum.

The structural width of the interfacial region was determined in two ways. First, visual inspection and measurement of the interface width were conducted from high magnification prints. Secondly, intensity profiles perpendicular to the interface were made from the recorded images. Measurements were taken from more than seven different basal α-Al_2O_3/Al interfaces, and the structural width was found to be 0.8 ± 0.2 nm, where the main source of error is the definition of the last (0006) plane of the alumina, immediately at the interface. The width of (0001) α-Al_2O_3/Al interfaces was measured for both types of sample and no difference was found in the width or periodicity of the interface region.

PEELS and EDS analyses were conducted on the same interfaces as investigated by HREM. This was accomplished by recording images of the region containing the interface at different magnifications (immediately after recording HREM images), which then served as a guide to locate the same interface in the STEM. Kikuchi patterns were used in the STEM to orient the alumina grain in the same zone axis as used for HREM, thus ensuring that the interface was parallel to the STEM electron beam. Figure 33 shows EDS spectra of the interface shown in

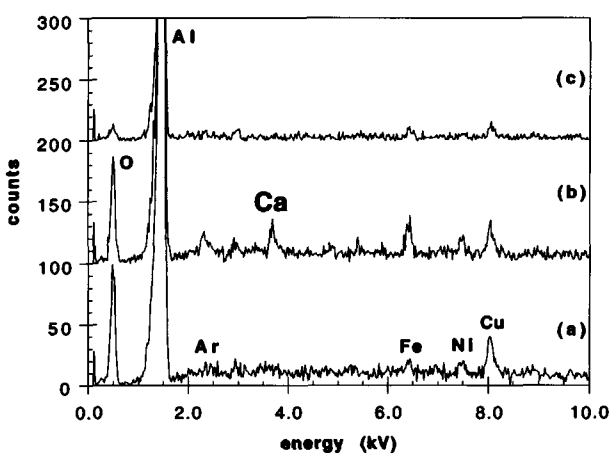


Figure 33. EDS spectra of the interface region shown in Fig. 32, (a) in the alumina, (b) directly at the interface, and (c) in the aluminum. The presence of a small oxygen peak in the aluminum spectrum is due to surface oxide. Argon, iron, nickel, and copper are confirmed artifacts from specimen thinning. Spectra (b) and (c) are shifted vertically for display purposes.

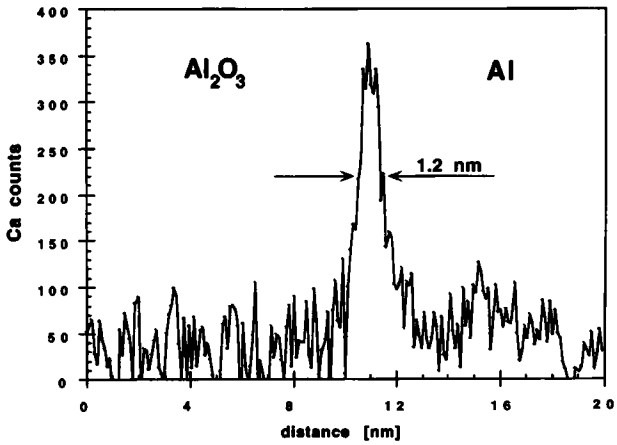


Figure 34. Calcium profile from a PEELS line scan perpendicular to an (0001) α-Al_2O_3/Al interface.

Fig. 32. The spectra were recorded by rastering the beam over a $10 \times 12\,nm^2$ area, (a) in the alumina, (b) directly at the interface, and (c) in the aluminum. The presence of calcium is clearly visible, with a peak significantly above the background level. Quantification of five interfaces by EDS resulted in an average calcium interface excess of $\Gamma = (2.0 \pm 1.0)\,Ca/nm^2$. PEELS line scans perpendicular to the interface were also made for five interfaces, and an example is shown in Fig. 34. Quantification of the calcium excess from PEELS line scans resulted in $\Gamma = (2.5 \pm 0.5)\,Ca/nm^2$. EDS and PEELS were conducted for (0001) α-Al_2O_3/Al interfaces for both types of sample; calcium segregation was found for both types of sample and no significant difference in the excess was detected.

The full-width-half-maximum (FWHM) of the calcium profile from the PEELS line scan provides an upper limit of calcium distribution perpendicular to the interface. This was measured to be $\sim 1.2\,nm$. This value is only an upper limit because it is a convolution of the real calcium distribution with the electron beam intensity profile and therefore agrees with the HREM results.

Figure 35 is a noise-filtered high magnification HREM micrograph of an α-Al_2O_3 basal twin boundary. The twin has been identified as a deformation twin, due to the presence of Moiré-type fringes located

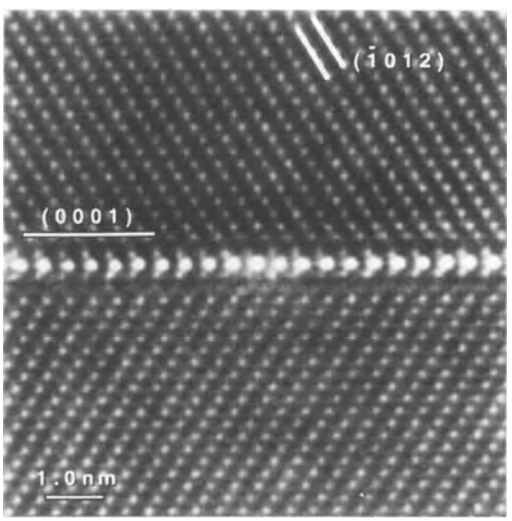


Figure 35. High magnification HRTEM image of a basal twin boundary with a calcium excess of $\Gamma = 1.0 \pm 0.5\,Ca/nm^2$. The zone axis of the upper side of the boundary is [1210]. The objective lens defocus and simulated specimen thickness, as determined from quantitative comparison to image simulations, are $-55\,nm$ and $8.6\,nm$, respectively. The image has not been filtered or processed in any manner.

along its length in the thicker regions of the specimen, which are not found in basal growth twins in α-Al_2O_3 [124]. While the twin boundary shows a pattern and contrast that nominally fit the cation-terminated basal twin structure [124], quantitative HREM image analysis to determine possible atom positions is currently in process. EDS of the twin boundary revealed a calcium excess of $\Gamma = 1.0 \pm 0.5\,Ca/nm^2$.

It should be noted that a detailed search for the presence of glass was conducted using both conventional and HRTEM. No glass phase was found at the interface regions, or at the general alumina grain boundaries or triple junctions. Furthermore, no silica segregation was detected at the interfaces or general grain boundaries, even though the starting powders contained a silica content comparable to that of the calcium. The reasons for this are still not clear.

From the combined results it can be concluded that calcium does segregate to basal surfaces in α-Al_2O_3. Assuming that the calcium cations substitute for the aluminum cations at the (0001) α-Al_2O_3/Al interfaces [125], the measured calcium excess correlates to a coverage (monolayer) of $\theta = 0.25$.

However, from the combined HREM and PEELS line scans, it is evident that the calcium is distributed over 4 ± 1 cation planes, which correspond to a Ca/Al concentration of approximately 8 at.% or a nominal composition correlating to the bulk phase $CaO.6\alpha$-Al_2O_3 [126]. While the detailed atomic structure of the surface phase is still not known, it may correlate to the Ca-driven (2×2) basal surface reconstruction [127]. The presence of the surface phase would be expected to alter the contact angle and wetting of aluminum on alumina and may explain the lack of wetting defects in the composite. Further structural and chemical analysis of the surface phase is currently in progress.

6.7.3 Fiber/Whisker Reinforcement

Practical ceramic matrix composites reinforced with continuous fibers exhibit important failure/damage behaviors in mode I, mode II, and mixed mode I/II as well as in compression. The failure sequence depends on whether the reinforcement is uniaxial or multiaxial and whether woven or laminated patterns are used. However, the underlying failure processes are illustrated by the behavior of uniaxially reinforced systems [24]. The composite's properties are known to be dominated by the interface, and bounds must be placed on the interface debonding and sliding resistance in order to have a composite with attractive mechanical properties. The strong dependence of ceramic matrix composite properties on the mechanical properties of the interface generally demands consideration of fiber coatings and/or reaction product layers. Residual stresses caused by thermal expansion differences are also very important.

The microscopy of interfaces and interphases plays a dominant role in the study of the ceramic composite. TEM is able to determine structure, composition, and chemical reactions at these interfaces. Possible chemical reactions may lead to a change in the interphase properties and therefore also of the composites. The possible chemical reaction may limit the applicability of fiber-reinforced ceramics.

Acknowledgement

The author is grateful to his colleagues Dr. F. Ernst, Dr. H.-J. Kleibe)Bayreuth), and Dr. J. Mayer for many stimulating discussions. The work was financially supported through the Volkswagen foundation under contracts I/70082 and I/70844 (Dr. H. Steinhardt, contract monitor).

6.8 References

[1] W. D. Kingery, H. K. Bowen, D. R. Uhlmann, *Introduction to Ceramics*, 2nd ed., Wiley, New York **1960**.

[2] G. F. Vander Voort, *Metallography, Principles and Practice*, McGraw-Hill, New York **1984**.

[3] L. Reimer, *Scanning Electron Microscopy*, Springer, Berlin **1985**.

[4] D. A. Bonnell (Ed.), *Scanning Tunneling Microscopy and Spectroscopy: Theory, Techniques, and Applications*, VCH, Weinheim **1993**.

[5] R. Wiesendanger, *Scanning Probe Microscopy and Spectroscopy: Methods and Applications*, Cambridge University Press, Cambridge **1994**.

[6] E. E. Underwood, *Quantitative Stereology*, Addison-Wesley, Reading, MA **1970**.

[7] A. P. Sutton, R. W. Balluffi, *Interfaces in Crystalline Materials*, Oxford University Press, Oxford **1995**.

[8] L.-C. Dufour, C. Monty, G. Petot-Ervas (Eds.), *Surfaces and Interfaces of Ceramic Materials* (NATO ASI Series, Series E: Applied Sciences, Vol. 173), Kluwer, Dordrecht, The Netherlands **1989**.

[9] R. M. Fulrath, J. A. Pask (Eds.), *Ceramic Microstructure '76: With Emphasis to Energy Related Applications*, Westview Press, Boulder, CO **1967**.

[10] R. M. Fulrath, J. A. Pask (Eds.), *Ceramic Microstructure '76: With Emphasis to Energy Related Applications*, Westview Press, Boulder, CO **1977**.

[11] J. A. Pask, A. G. Evans (Eds.), *Ceramic Microstructure '86: Role of Interfaces*, Plenum, London **1987**.

[12] M. Swain (Ed.), *Structure and Properties of Ceramics in Materials Science and Technology, Vol. 11*, VCH, Weinheim **1994**.

[13] G. Petzow, *Metallographisches, Keramographisches, Plastographisches Ätzen*, 6th ed., Gebrüder Bornträger, Berlin **1994**.

[14] B. Baretzky, B. Reinsch, U. Täffner, G. Schneider, M. Rühle, *Z. Metallkd.* **1996,** *87*, 332.

[15] P. Hirsch, A. Howie, R. B. Nicholson, D. W. Pashley, M. J. Whelan, *Electron Microscopy of Thin Crystals*, 2nd ed., Robert E. Krieger, New York **1977**.

[16] H. Bethge, J. Heydenreich (Eds.), *Electron Microscopy in Solid State Physics* (Materials Science Monographs: 40), Elsevier, Amsterdam **1987**.

[17] J. C. H. Spence, *Experimental High-Resolution Electron Microscopy*, 2nd ed., Oxford University Press, Oxford **1988**.

[18] S. Horiuchi, *Fundamentals of High-Resolution Transmission Electron Microscopy*, Elsevier, Amsterdam **1994**.

[19] D. C. Joy, A. R. Romig, Jr., J. I. Goldstein (Eds.), *Principles of Analytical Electron Microscopy*, Plenum, New York **1986**.

[20] R. F. Egerton, *Electron Energy-Loss Spectroscopy in the Electron Microscope*, 2nd ed., Plenum, New York **1996**.

[21] J. Bruley, *Microsc. Microanal. Microstr.* **1993,** *4*, 23.

[22] R. Brydson, H. Sauer, W. Engel, J. M. Thomas, E. Zentler, *J. Chem. Soc., Chem. Commun.* **1989,** *15*, 1010.

[23] E. Ryshkewitch, D. W. Richerson, *Oxide Ceramics*, 2nd ed., Academic, Orlando, FL **1986**.

[24] M. Rühle, A. G. Evans, *Prog. Mater. Sci.* **1989,** *33*, 85.

[25] D. J. Green, R. H. J. Hannink, M. V. Swain, *Transformation Toughening of Ceramics*, CRC Press, Boca Raton, FL **1989**.

[26] A. G. Evans, *J. Am. Ceram. Soc.* **1990,** *73*, 187.

[27] J. D. Cawley, W. E. Lee in *Materials Science and Technology* (Eds.: R. W. Cahn, P. Haasen, E. J. Kramer), *Vol. 11, Structure and Properties of Ceramics* (Ed.: M. V. Swarz), VCH, Weinheim **1994,** p. 47.

[28] W. H. Gibson (Ed.), *Alumina as a Ceramic Material*, The American Ceramic Soc., Columbus, OH **1970**.

[29] E. Dörre, H. Hübner, *Alumina*, Springer, Berlin **1984**.

[30] W. E. Lee, W. M. Rainforth, *Ceramic Microstructures*, Chapman and Hall, London **1994**.

[31] S. J. Wilson, *Br. Ceram. Soc. Proc.* **1979,** *28*, 281.

[32] H. Tanaka, *Seimitsu Kikai* **1973,** *39*, 917.

[33] Y. Yamamoto in *Silicon Carbide, Silicon Nitride and Fibre Reinforced Ceramics* (Ed.: Y. Yamamoto), **1992,** TLC, Osaka, p. 399.

[34] P. F. Becher, G. C. Wei, *J. Am. Ceram. Soc.* **1984,** *67*, C-267.

[35] G. C. Wei, P. F. Becher, *Am. Ceram. Soc. Bull.* **1985,** *64*, 208.

[36] N. Claussen, J. Steeb, R. F. Pabst, *Am. Ceram. Soc. Bull.* **1977,** *56*, 559.
[37] W. D. Kingery (Ed.), *Advances in Ceramics Vol. 10*, The American Ceramic Soc., Columbus, OH **1984.**
[38] A. G. Evans, Y. Fu, *Advances in Ceramics 10*, The American Ceramic Soc., Columbus, OH **1984,** p. 697.
[39] T. Höche, P. R. Kenway, H. J. Kleebe, M. Rühle, *J. Am. Ceram. Soc.* **1994,** *77*, 339.
[40] T. Höche, P. R. Kenway, H. J. Kleebe, M. W. Finnis, M. Rühle, *J. Phys. Chem. Solids* **1994,** *55*, 1067.
[41] J. Bruley, T. Höche, H.-J. Kleebe, M. Rühle, *J. Am. Ceram. Soc.* **1994,** *77*, 2273.
[42] W. D. Kaplan, H. Müllejans, M. Rühle, J. Rödel, N. Claussen, *J. Am. Ceram. Soc.* **1995,** *78*, 2841.
[43] T. Höche, M. Rühle, *J. Am. Ceram. Soc.* **1996,** *79*, 1961.
[44] H. Grimner, R. Bounet, S. Lartigue, L. Priester, *Phil. Mag.* **1990,** *61*, 493.
[45] D. Bouchet, F. Dupar, S. Lartigue-Korinek, *Microsc. Microanal. Microstr.* **1993,** *4*, 561.
[46] W. Swiatuicki, S. Lartigue-Korinek, A. Dubon, J. Y. Laval, *Mater. Sci. Forum* **1993,** *126–128*, 193.
[47] M. Rühle, *Fresenius Z. Anal. Chem.* **1994,** *349*, 49.
[48] M. Rühle, G. Necker, W. Mader, *Ceram. Trans.* **1989,** *5*, 340.
[49] H. F. Fischmeister, G. Elssner, B. Gibbesch, K.-H. Kadow, D. Korn, F. Kawa, W. Mader, M. Turwitt, *Rev. Sci. Instrum.* **1992,** *64*, 234.
[50] P. R. Kenway, *Phil. Mag.* **1993,** *B68*, 171.
[51] E. Schumann, J. C. Yang, M. Rühle, M. J. Graham, *MRS Symp. Proc. Symp. Proc.* **1995,** *364*, 1291.
[52] J. Fang, A. Marck Thompson, M. P. Harmer, H. M. Chan, *J. Am. Ceram. Soc.* **1996,** *79*, in print.
[53] S. J. Bennison, M. P. Harmer, *Ceram. Trans.* **1990,** *7*, 13.
[54] S. I. Bae, S. Baik, *J. Am. Ceram. Soc.* **1994,** *77*, 2499.
[55] R. C. Garvie in *High Temperature Oxides, Part II* (Ed.: A. M. Alper), Academic, New York **1970,** p. 117.
[56] A. H. Heuer, L. W. Hobbs (Eds.), *Science and Technology of Zirconia, Advances in Ceramics Vol. 3*, The American Ceramic Society, Columbus, OH **1981.**
[57] N. Claussen, M. Rühle, A. H. Heuer (Eds.), *Science and Technology of Zirconia II, Advances in Ceramics Vol. 12*, The American Ceramic Society, Columbus, OH **1984.**
[58] S. Somiya, N. Yamamoto, H. Hanagida (Eds.), *Science and Technology of Zirconia III, Advances in Ceramics Vol. 24*, The American Ceramic Society, Columbus, OH **1988.**
[59] R. H. J. Hannink, *Science and Technology of Zirconia IV*, Amsterdam **1992.**
[60] E. C. Subbarao, *Adv. Ceram.* **1981,** *3*, 1.
[61] G. K. Bansal, A. H. Heuer, *Acta Metall.* **1972,** *20*, 1281.
[62] G. K. Bansal, A. H. Heuer, *Acta Metall.* **1974,** *22*, 409.
[63] R. C. Garvie, R. H. J. Hannink, R. T. Pascoe, *Nature (London)* **1975,** *258*, 703.
[64] A. G. Evans, R. M. Cannon, *Acta Metall.* **1986,** *34*, 761.
[65] N. Claussen in *Science and Technology of Zirconia II, Advances in Ceramics Vol. 12* (Eds.: N. Claussen, M. Rühle, A. H. Heuer), The American Ceramics Society, Columbus, OH **1984,** p. 325.
[66] S. Schmauder, W. Mader, M. Rühle in *Science and Technology of Zirconia II, Advances in Ceramics Vol. 12* (Eds.: N. Claussen, M. Rühle, A. H. Heuer), The American Ceramic Society, Columbus, OH **1984,** p. 25.
[67] N. Claussen in *Science and Technology of Zirconia II, Advances in Ceramics Vol. 12* (Eds.: N. Claussen, M. Rühle, A. H. Heuer), The American Ceramic Society, Columbus, OH **1984,** p. 325.
[68] V. Lanteri, T. E. Mitchell, A. H. Heuer in *Science and Technology of Zirconia II, Advances in Ceramics Vol. 12* (Eds.: N. Claussen, M. Rühle, A. H. Heuer), The American Ceramic Society, Columbus, OH **1984,** p. 118.
[69] W. M. Kriven, W. L. Fraser, S. W. Kennedy in *Science and Technology of Zirconia, Advances in Ceramics Vol. 3* (Eds.: A. H. Heuer, L. W. Hobbs), The American Ceramic Society, Columbus, OH **1981,** p. 82.
[70] M. Rühle, W. M. Kriven in *Proc. Int. Conf. on Solid-to-Solid Phase Transformation* (Eds.: H. I. Aaronson, D. E. Laughlin, R. F. Sekerka, C. M. Wayman), AIME, Pittsburgh, PA **1982,** p. 1569.
[71] M. Rühle, W. M. Kriven, *Ber. Bunsenges. Phys. Chem.* **1983,** *87*, 222.
[72] W. Mader, M. Rühle in *Proc. 10th Int. Congress on Electron Microscopy, Vol. 2*, **1982,** p. 103.
[73] W. Mader, *Phil. Mag.* **1987,** *A55*, 59.
[74] M. Rühle, N. Claussen, A. H. Heuer in *Science and Technology of Zirconia II, Advances in Ceramics Vol. 12* (Eds.: N. Claussen, M. Rühle, A. H. Heuer), The American Ceramic Society, Columbus, OH **1984,** p. 352.
[75] H. Schubert, *J. Am. Ceram. Soc.* **1986,** *69*, 270.

[76] M. Rühle, B. Kraus, A. Strecker, D. Waidelich in *Science and Technology of Zirconia II, Advances in Ceramics Vol. 12* (Eds.: N. Claussen, M. Rühle, A. H. Heuer), The American Ceramic Society, Columbus, OH **1984,** p. 256.
[77] M. Rühle, L. T. Ma, W. Wunderlich, A. G. Evans, *Physica* **1988,** *B150*, 86.
[78] M. Rühle, A. H. Heuer in *Science and Technology of Zirconia II, Advances in Ceramics Vol. 12* (Eds.: N. Claussen, M. Rühle, A. H. Heuer), The American Ceramic Society, Columbus, OH **1984,** p. 284.
[79] I.-W. Chen, Y.-H. Chiao, *Acta Metall.* **1983,** *31*, 1627.
[80] I.-W. Chen, Y.-H. Chiao, *Acta Metall.* **1985,** *33*, 1827.
[81] A. G. Khachaturyan, *Theory of Structural Transformations in Solids*, Wiley, New York **1983.**
[82] V. Lanteri, T. E. Mitchell, A. H. Heuer, *J. Am. Ceram. Soc.* **1983,** *66*, 23.
[83] T. Mura, *Micromechanics of Defects*, Martinus Nijhoff, The Hague **1982.**
[84] F. E. Buresch, *Fract. Mech. Ceram.* **1975,** *4*, 835.
[85] R. G. Hoagland, J. D. Embury, D. J. Green, *Scr. Metall.* **1975,** *9*, 907.
[86] A. G. Evans, *Scr. Metall.* **1976,** *10*, 93.
[87] M. Rühle, N. Claussen, A. H. Heuer, *J. Am. Ceram. Soc.* **1986,** *69*, 195.
[88] M. Rühle, A. G. Evans, R. M. McMeeking, J. W. Hutchinson, *Acta Metall.* **1987,** *35*, 2701.
[89] A. G. Evans, K. T. Faber, *J. Am. Ceram. Soc.* **1984,** *67*, 255.
[90] J. W. Hutchinson, *Acta Metall.* **1987,** *35*, 1605.
[91] D. R. Clarke, *Ultramicroscopy* **1979,** *4*, 33.
[92] W. K. Wadhawa, *Phase Trans.* **1982,** *3*, 3.
[93] J. F. Jue, A. V. Virkar, *J. Am. Ceram. Soc.* **1990,** *73*, 3650.
[94] C. J. Chan, F. F. Lange, M. Rühle, J. F. Jue, A. V. Virkar, *J. Am. Ceram. Soc.* **1991,** *74*, 807.
[95] A. Foitzik, M. Stadtwald-Klenke, M. Rühle, *Z. Metallkd.* **1993,** *84*, 6.
[96] D.-S. Cheong, A. Dominguez-Rodriguez, A. H. Heuer, *Phil. Mag.* **1989,** *A60*, 123.
[97] J. Martinez-Fernandez, M. Jimenez-Melendo, A. Dominguez-Rodriguez, A. H. Heuer, *J. Am. Ceram. Soc.* **1990,** *23*, 2452.
[98] D.-S. Cheong, A. Dominguez-Rodriguez, A. H. Heuer, *Phil. Mag.* **1991,** *A63*, 377.
[99] B. Baufeld, M. Bartsch, U. Messerschmidt, D. Baiker, *Acta Metall. Mater.* **1995,** *43*, 1925.
[100] D. Baiker, B. Baufeld, U. Messerschmidt, *Phys. Status Solidi (a)* **1993,** *137*, 569.
[101] F. L. Riley (Ed.), *Nitrogen Ceramics*, Noordhofer, Leyden **1977.**
[102] F. L. Riley (Ed.), *Progress in Nitrogen Ceramics*, Martinus Nijhoff, The Hague **1983.**
[103] D. A. Bonnell, T. Y. Tien (Eds.), *Preparation and Properties of Silicon Nitride Based Materials, Vol. 47*, Materials Science Forum **1989.**
[104] I.-W. Chen, P.-T. Becher, M. Mitomi, G. Petzow, T.-S. Yen, *Silicon Nitride Ceramics—Scientific and Technological Advances, Vol. 287*, MRS Symp. Proc. Symp. Proc. **1993.**
[105] M. J. Hoffmann, G. Petzow, *Tailoring of Mechanical Properties of Si_3N_4 Ceramics*, NATO ASI Series, Series E: Applied Sciences Vol. 276, Kluwer, Dordrecht, The Netherlands **1994.**
[106] M. Peukert, D. Greil, *J. Mater. Sci.* **1987,** *22*, 213.
[107] J. Weiss, W. A. Kaysser in *Progress in Nitrogen Ceramics* (Ed.: F. F. Riley), Martinus Nijhoff, The Hague **1983.**
[108] F. F. Lange, *J. Am. Ceram. Soc.* **1979,** *62*, 428.
[109] M. Mitomo, M. Tsutsumi, H. Tanaka, S. Uenosono, F. Saito, *J. Am. Ceram. Soc.* **1990,** *73*, 2441.
[110] M. J. Hoffmann, G. Petzow, *MRS Symp. Proc. Symp. Proc.* **1993,** *287*, 3.
[111] P. F. Becher, H. T. Lin, S. L. Hwang, M. J. Hoffmann, I.-W. Chen, *MRS Symp. Proc. Symp. Proc.* **1993,** *287*, 147.
[112] D. R. Clarke, *J. Am. Ceram. Soc.* **1987,** *70*, 15.
[113] H. J. Kleebe, M. K. Cinibulk, R. M. Cannon, M. Rühle, *J. Am. Ceram. Soc.* **1993,** *76*, 1969.
[114] I. Tanaka, H.-J. Kleebe, M. K. Cinibulk, J. Bruley, D. R. Clarke, M. Rühle, *J. Am. Ceram. Soc.* **1994,** *77*, 911.
[115] X. Pan, H. Gu, R. van Weeren, S. C. Danforth, R. Cannon, M. Rühle, *J. Am. Ceram. Soc.* **1996,** *79*, in print.
[116] H.-J. Kleebe, M. K. Cinibulk, I. Tanaka, J. Bruley, R. M. Cannon, D. R. Clarke, M. J. Hoffmann, *MRS Symp. Proc. Symp. Proc.* **1993,** *287*, 65.
[117] H.-J. Kleebe, J. Bruley, M. Rühle, *J. Europ. Ceram. Soc.* **1994,** *12*, 1.
[118] I. Tanaka, G. Pezzotti, T. Okamoto, Y. Miyamoto, M. Koizumi, *J. Am. Ceram. Soc.* **1989,** *72*, 1656.
[119] J. Bruley, I. Tanaka, H.-J. Kleebe, M. Rühle, *Anal. Chim. Acta* **1994,** *297*, 97.
[120] D. R. Clarke, T. M. Shaw, A. P. Philipsi, R. G. Horn, *J. Am. Ceram. Soc.* **1993,** *76*, 1201.
[121] H. Prielipp, M. Knechtel, N. Claussen, S. K. Streiffer, H. Müllejans, M. Rühle, J. Rödel, *Mater. Sci. Eng.* **1995,** *A197*, 19.
[122] R. C. McCune, R. C. Ku, *Adv. Ceram.* **1984,** *10*, 217.

[123] S. Baik, C. L. White, *J. Am. Ceram. Soc.* **1987,** *70*, 682.
[124] W. D. Kaplan, P. R. Kenway, D. G. Brandon, *Acta Metall. Mater.* **1995,** *43*, 835.
[125] R. W. Grimes, *J. Am. Ceram. Soc.* **1994,** *77*, 378.
[126] A. Utsunomiya, K. Tanaka, H. Morikawa, F. Maruko, *J. Solid State Chem.* **1988,** *75*, 197.
[127] M. Gautier, G. Renaud, L. P. Van, B. Villette, M. Pollak, N. Thromat, F. Jollet, J.-P. Duraud, *J. Am. Ceram. Soc.* **1994,** *77*, 323.

Handbook of Microscopy
S. Amelinckx, D. van Dyck,J. van Landuyt , G. van Tendeloo

7 Microscopy of Gemmological Materials

"A material, to be considered a gem, must have beauty, durability and scarcity". This is the definition as listed in the reference textbooks [1, 2]. Gemmology is the study of these materials.

7.1 Optical Microscopy

7.1.1 Introduction

Today the optical microscope is one of the most important instruments of the gemmologist, independent of whether his interest is purely academic or of a more commercial nature. Upon its introduction in the field of gemstones at the beginning of this century, the gemmological microscope hardly differed from the optical microscopes that were commonly used in biology or petrography [3]. It should be borne in mind that these are somewhat special versions of the optical microscope, described in detail in Chap. 1 of Vol. 1 of this Handbook, to which we refer for the physical background on instrument and image interpretation.

Since most of the pioneers of gemmology had a geological or mineralogical background, the petrographical microscope became the standard instrument for the study of gems. The introduction of the microscope in gemmology, and in fact the founding of modern gemmology, came about with the successful introduction, on the gemstone market around the turn of the century, of important quantities of artificial ('synthetic') high quality rubies. These had been produced by means of the patented (1905) 'Verneuil' flame-fusion technique. To avoid the trade of natural, especially Burmese, rubies being disrupted or even coming to a complete stop, accurate and reliable identification techniques became necessary. Soon it was realized that an accurate identification technique consisted of observing the internal features of the gemstone. If synthetic, the latter would reveal bubbles and curved growth lines that could not be mistaken for inclusions in natural stones. This required, of course, the use of a magnifying instrument, the logical choice being a microscope.

In the years that followed, the gemmological microscope gradually evolved into a more specialized instrument. This evolution took two different routes. The rules for grading diamonds—the famous four Cs—are completely different from those determining the value of other precious stones. Two types of microscope were developed accordingly.

7.1.2 Properties and Use of the Gemmological Microscope

In the study of precious stones other than diamonds[1], the microscope is mostly used to identify the gem rather than for grading purposes, as is usually the case with diamonds. The most important questions to be solved by the gemmologist are:

1. What material is it?
2. Is the material natural?
3. Are the properties of this material natural or have they been artificially induced or enhanced?

Commercially, there is also a fourth question worth posing:

4. What is the geographical origin of the gemstone? As, for example, Colombian emeralds fetch better prices than the same quality Brazilian material.

Not all of these questions can be answered by microscopical observations. Other gemmological instruments are needed, such as the polariscope, the refractometer, the dichroscope, and the spectroscope. Not only the optical properties need proper recording: other physical parameters such as the specific gravity and thermal properties can also have a diagnostic value. The main difference between these two groups of properties is, however, the fact that most optical properties can be observed by means of the microscope with proper accessories. The identification of synthetic gems and the detection of treatments is mostly based on microscopical observation.

[1] Commercially speaking, these gems are called 'colorstones'.

The study of inclusions can be particularly rewarding, as pointed out by Gübelin [4]. The identification of the 'paragenesis', i.e., the inclusions that crystallized synchronously with the host, in a particular geological or chemical environment, allows the gemmologist not only to conclude whether the gem is natural or synthetic, but in certain limited cases it is also possible to deduce from which geographical origin or geological deposit it was extracted [5]. Certain treatments, such as heat treatment, also have a marked effect on the gem and its inclusion content, as can be revealed by microscopical observation [6].

7.1.2.1 Stereoscopic Capability

As is the case with 'diamond microscopes', most 'gemmological microscopes' are really stereoscopic microscopes. Stereoscopic microscopes differ from standard or biological microscopes and from polarizing or petrographic microscopes having split optical paths. The positioning of the separate optical paths at angles between 12 and 16° produces a pair of images (Fig. 1) similar to those produced by human eyes when viewing any object at a ± 25 cm distance, which enables 3D vision.

Obviously this type of microscope is ideal for viewing nonflat objects; it was mainly developed for the electronics industry where the manipulation of small components is a key task. Since many electronic components and gems are more or less similar in size, the use of the same optics for both kinds of material is only natural. To benefit from the 3D capability of this type of optics, oblique or reflective illumination in the case of

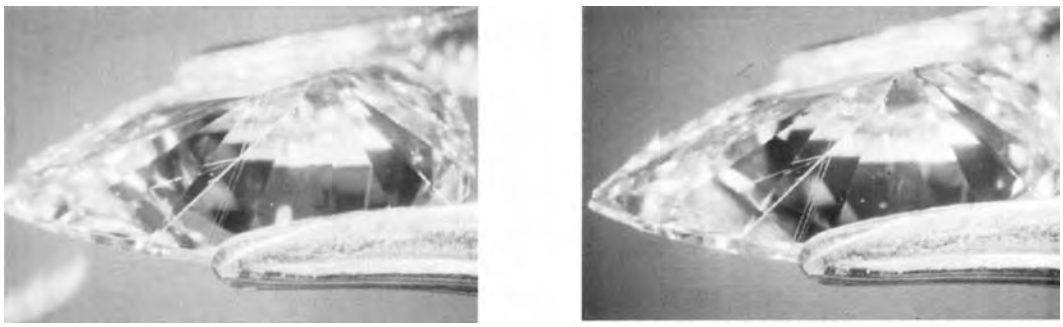


Figure 1. Photomicrographs of a pear-shaped diamond as seen through a stereoscopic microscope. The left and right images [(a) and (b), respectively], showing a pear-shaped diamond with some structural features (surface growth lines) and an associated needle-like inclusion, are clearly different. Together they produce a stereoscopic effect similar to normal binocular viewing. The ability to see in 3D is very important to the gemmologist in order to determine whether the observed features are on the surface (external) or underneath the surface (internal). An important part of the quality evaluation of diamonds is based on this difference.

transparent, translucent, and opaque materials, and dark field illumination for transparent materials, are compulsory. Independent of the quality of the optics, the maximum magnification of stereoscopic microscopes is mostly limited to approximately 100×. Nevertheless, accessory lenses exist that make it possible to increase or decrease the total magnification, e.g. a 0.5× lens can be very useful when examining large gems macroscopically[2].

Two major types of stereoscopic microscopes can be distinguished [7]:

1. The Galilei-type stereoscopic microscope consists of two separate oculars which are served by only one shared objective (Fig. 2a).
2. The Greenough-type stereoscopic microscope consists of two separate oculars which are served by two fully separate objectives (Fig. 2b).

Both types can be equipped with sets of objectives with fixed magnifications on a revolver, or with a zoom lens allowing any magnification within a given range. The advantages of the fixed objectives are accuracy of magnification and better resolution. The main advantage of a zoom is obviously the flexibility to choose the optimal magnification most suitable for the job.

7.1.2.2 Depth of Field

Very important to the gemmologist is the depth of field of the microscope. Proper understanding and careful application of this property can be of considerable help in identifying small features on or in the stone. This is especially important in diamond grading where the presence or absence of inclusions is an important quality characteristic. The depth of view or focus depends on the objective magnification and on the aperture size. It decreases at higher magnifications. For the Bausch & Lomb SZ 7, a stereoscopic microscope[3] which is very popular in

[2] Increasing the magnification is of course limited by the resolving power of the objectives.

[3] Today the Bausch & Lomb SZ 7 is manufactured by Leica.

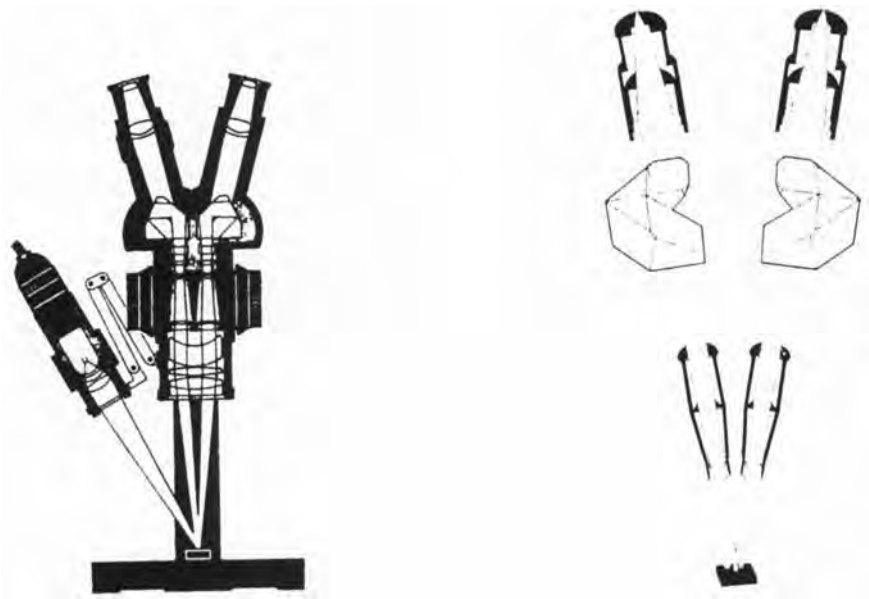


Figure 2. Comparison of two important types of stereoscopic microscope (modified from Van Maaren [7]). (a) Diagram showing the light path in a Galilei-type stereoscopic microscope. Characteristic is the single objective lens shared by the two oculars. (b) Diagram showing the light path in a Greenough-type stereoscopic microscope. Each ocular is served by a different objective lens.

diamond grading, the depth of field ranges from 1 mm at its lowest magnification (1×) to 0.05 mm at the highest magnification (7×).

This means that for this particular type of microscope the position of a small inclusion at an objective magnification of 1× can only be assessed within a 1 mm thick depth range. At the highest objective magnification (7× for this particular microscope), the inclusion can be traced within a layer of only 0.05 mm thickness.

7.1.2.3 Illumination

In diamond grading only two distinct types of (standardized) illumination, i.e. dark field and reflective illumination, are customary. A gemmological microscope must have more illumination possibilities in order to be able to observe all the characteristic features that allow identification of the gem under scrutiny. Transmitted illumination is considered as standard in gem identification. Simple transmitted illumination is perfectly suited for flat (plane-parallel) objects, but is completely inadequate when used with highly refractive, polished gems. This is mainly because total internal reflection on the facets prevents light reaching the eye of the observer, and as a consequence the gem shows dark against a blinding background, thus obscuring the gem for inspection. To arrive at an ideal transmitted illumination, the gem must therefore be placed in an immersion cell in a liquid having approximately the same refractive index as that of the gem (Table 1). This transmitted–immersion lighting is especially effective at bringing about subtle

Table 1. Abbreviated list of immersion liquids and their refractive indices, as used in the transmitted–immersion microscopy of gems (in order of increasing RI)[a]

Liquid	RI	Liquid	RI	Liquid	RI
Water	1.33	Cedar oil	1.51	Cassia oil	1.60
Alcohol	1.36	Monochlorobenzene	1.53	Monoiodobenzene	1.62
Amyl acetate	1.37	Canada balsam	1.53	Monochloronaphthalene	1.63
Petroleum	1.45	Clove oil	1.54	Monobromonaphthalene	1.66
Carbon tetrachloride	1.46	Nitrobenzene	1.55	Monoiodonaphthalene	1.70
Glycerin	1.47	Benzylbenzoate	1.56	Methylene iodide	1.74
Castor oil	1.48	Orthotoluidine	1.57	Methylene iodide + S	1.78
Toluene	1.49	Aniline	1.58	Methylene iodide + S + C_2I_4	1.81
Benzene	1.50	Bromoform	1.59	West's solution	2.05

[a] Adapted after Webster [1].

color zoning and other structural features, such as the characteristic curved growth lines in Verneuil (flame-fusion)-type synthetic rubies.

A side effect of the transmitted–immersion illumination technique is that the stereoscopic viewing is strongly compromised. Obtaining an optimal illumination level simultaneously in both oculars is almost impossible. This effect is even more evident when using polarizing filters. It can therefore be argued that when using only transmitted–immersion illumination, a monotubular microscope would probably be more effective.

Table 1 shows that immersion liquids, which are neither too costly nor too toxic, can be obtained up to a refractive index (RI) of approximately 2. For higher refractive indices, special liquids have to be prepared, which are mostly toxic and rather unstable. To obtain a liquid with an RI of approximately 2.415 (the RI of diamond), a glasslike substance (a complex borate containing lead and bismuth) would have to be used and melted. These practical constraints limit the use of immersion microscopy to liquids having an RI of ±1.75. In transmitted illumination, the observer is looking straight into the parallel light beam, resulting in a low visibility of whitish inclusions. In dark field illumination, on the other hand, the beams of light converge on the gem and are not seen as such, leaving the interior of the gem unlit. Only inclusions, surface features (dust, blemishes, etc.) and facet edges will stand out brightly on a dark background (Fig. 3).

Two important consequences of this kind of illumination are:

1. In order to avoid seeing too much dust and fingerprints on the gem, the stone has to be cleaned prior to the examination.
2. Reflections of light beams on facet edges will be seen, and have to be identified as such. This requires additional skills of the gemmologist.

Another type of illumination, which is restricted to certain applications in gemmology, is so-called coaxial illumination. In coaxial illumination, a narrow, focused, intense beam of light is directed to the object, from a position as close as possible to the objective and parallel to the latter's optical axis. This produces a well-lit area on the object's surface, precisely positioned in the center of the image of the

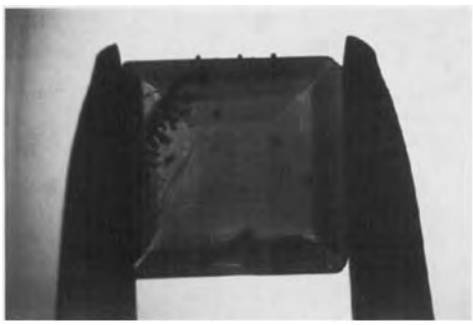


Figure 3. Comparison of transmitted–immersion illumination with dark field illumination. (a) Polished ruby in transmitted–immersion illumination. Only the inclusions are clearly visible; the facet edges are almost completely invisible. Small color zonations are very clearly visible. View perpendicular to the table; 15×. (b) A polished diamond as seen through a typical diamond microscope equipped with dark field illumination. The numerous reflections of the different facets and inclusions necessitate a good understanding by the observer of the behavior of light in a transparent, multi-faceted environment. View perpendicular to the pavilion facets; 15×.

objective, which is almost glare-free, thereby dramatically enhancing the visibility of any surface features which might be present. Typical applications are the study of the surface of pearls and the identification of synthetic opals and heat-treated corundum, to name but a few.

Whereas in research microscopes (certain types of metal microscopes and reverse biological microscopes) coaxial lighting can be obtained using special objectives with beam splitters, in gemmology coaxial illumination is mostly obtained using small microscopy lamps which are attached to the microscope's optics and directed at the object; fiber optic light guides are also used for this purpose. Sometimes it is necessary to combine different types of illumination in order to be able to identify the inclusions or other characteristics under scrutiny. A very useful combination could consist of dark field illumination combined with reflected illumination. The dark field illumination would be used for general illumination and clearly displaying the inclusions, the added reflected illumination could, for instance, separate an opaque inclusion from a highly reflective transparent inclusion, which in transmitted light would produce a similar optical effect. An overview of the different types of illumination and their advantages in observing the different kinds of inclusion[4] is given in Table 2. From this table it can be concluded that for most types of inclusion, dark field illumination with or without reflection illumination is extremely effective. Only structural phenomena such as growth lines and color zonations will be difficult to observe. These types of inclusions are most effectively observed using transmitted–immersion illumination. It is also clear that transmitted–immersion illumination is the most versatile all-round technique available.

7.1.2.4 Polarizing Microscopy

In analogy with polarized microscopy as applied in petrography, the use of

[4] In this context only a descriptive, morphological classification is assumed.

Table 2. Comparison of the visibility of different morphological types of inclusion when illuminated with different types of illumination.

	Transmitted–immersion illumination (TR–IMM)	Dark field illumination (DF)	Reflected illumination (R)	Coaxial illumination (CA)
Crystal inclusions	xxxx[b]	xxxxx[a]	xxx[c]	x[e]
Negative crystals (cavities–voids)	xxxx	xxxxx	xxx	x
Fractures and cleavages	xxxx	xxxxx	xxx	xxx
Healed fractures and fingerprint inclusions	xxxx	xxxxx	xxx	x
Growth lines	xxxxx	xx[d]	x	x
Color zonation	xxxxx	x	x	x

[a] xxxxx: extremely visible; [b] xxxx: very visible; [c] xxx: clearly visible; [d] xx: visible; [e] x: poorly visible.

polarization filters can greatly enhance the identification capability of the gemmological microscope. Just as with the polarizing microscope or the polariscope, the illumination should ideally be parallel transmitted light. In dark field or reflected illumination, polarization is mostly incomplete. On a gemmological microscope, polarization is best obtained using transmitted–immersion illumination. Two different operational modes are possible:

1. *Orthoscopy* (the polarization directions of both polarizer and analyzer are crossed): This technique is especially helpful to discriminate between different types of inclusion based on their optical properties[5]. Otherwise almost invisible birefringent inclusions which are oriented differently to the host become very clearly visible. Strain analysis also becomes possible, and strain centers can readily be identified (Fig. 4). This can be improved further by adding optical compensator plates such as a first order red compensator [8].
2. *Conoscopy* (the polarizer and analyzer are perpendicularly oriented and an auxiliary lens comparable to the Amici–Bertrand lens of a petrographical microscope is inserted): This produces interference figures in birefringent gems and is helpful in discriminating between uniaxial and biaxial materials. This mode can usually only be applied on a monocular microscope. Webster [1] mentions the use of different types of immersion cells which were specifically designed to allow easy identification of the interference pattern of a gem. Among these are the Gübelin hollow glass sphere and the more elaborate Klein and Schlossmacher's immersion sphere fit onto a three-circle Leitz universal stage. Although yielding interesting information, these optical devices, which are not commercially available, have practically no usage today. It is, however, interesting to note that a separate instrument (not a microscope), the Figueras '93 Conoscope [9], is built using a hollow glass bulb much like the Gübelin glass sphere.

[5] Unlike minerals in thin sections, inclusions in gems can have very variable, undetermined dimensions which make identification based on Michel–Lévy charts impossible.

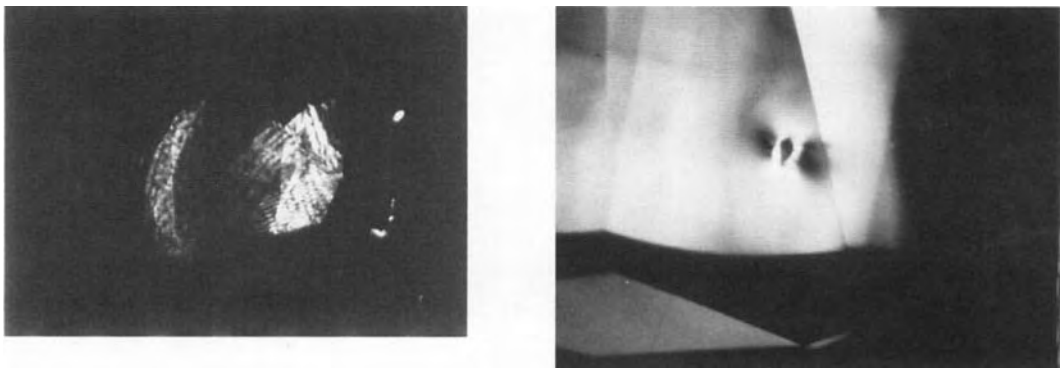


Figure 4. Effect of the application of polarization filters in gemstone microscopy. (a) The inclusion in the middle of this photomicrograph appears to be a strain center in this polished diamond. The contrast of the 'strain wings' around the inclusion is indicative of the amount of strain built up around it. Crossed polarizers; dark field illumination; 35×. (b) The presence of this characteristic interference pattern in an amethyst viewed in the direction of its optic axis is proof of the natural origin of the latter. The distinction between natural and synthetic amethysts is based on the typical Brazil twinning that is present in all natural amethysts and is virtually unknown in synthetic amethysts. Crossed polarizers; transmitted–immersion illumination; 10×. (Crowningshield et al. [23].)

7.1.2.5 Additional Illumination Techniques

Shadowing technique

Koivula [10] has effectively demonstrated that applying different stoppers in the beam of light in between the lamp and the gem produces a marked increase in the visibility of features which are otherwise difficult to discern. This effect could be compared to phase contrast microscopy. Phase contrast microscopy in gemmology was first introduced by Gübelin [11] in collaboration with the German microscope manufacturer Carl Zeiss. The high cost of this equipment and its limited applicability prevented a breakthrough.

Pinpoint illumination technique

A pinpoint illuminator originating from microsurgery has been adapted and modified by Koivula [12] for use in gem photomicrography. It consists of a highly flexible glass fiber light guide with interchangeable light walls of various shapes and sizes, the smallest only measuring approximately 1 mm in diameter. This illumination tool is especially useful for illuminating inclusions located in areas that are difficult to reach. It avoids glare and (sometimes) unwanted internal and external reflections. It is also very effective at showing thin film reflections such as those encountered in healing fractures.

7.1.2.6 Horizontal microscope

One of the constraints of the 'normal' microscope (i.e., a microscope with a vertically oriented light path) is that the stone holder partially obliterates the view of the stone when immersed in an immersion cell. On top of that, working bent over foul-smelling immersion liquids which are being heated by the lamp underneath is not exactly ideal. Not to mention the toxicity

of most of these liquids [13][6]. By 1951 Schlossmacher of Idar-Oberstein (Germany) had solved this problem by designing a new line of gemmological microscopes in which the optical path was horizontally oriented instead of vertically. The gem is now placed in a flat-faced squarish glass cuvette or immersion cell and the stone-holder's axis is vertically oriented at right angles to the optical path. This ensures an almost perfect view of the stone and at the same time a much improved maneuverability.

The committed use of the horizontal immersion microscope became, over the years, the trademark of German gemmology [14]. One of its characteristics is that most of the time the microscopical observation of any gem other than diamond is conducted in transmitted–immersion illumination. Dark field illumination and other types of illumination play only a modest role. This is in sharp contrast to the 'American'-oriented gemmology, or GIA 'gemology', which is based on the use of the famous 'Gemolite' microscope where dark field and reflected illumination are most frequently applied.

7.1.2.7 Further Applications

The Gemmological Microscope as a Refractometer: the Duke of Chaulnes Technique

The refractive index (RI) of any gem can be ascertained by means of the microscope by precisely measuring the displacement of the optical pod on focusing on the top and the bottom of the stone. This optical or apparent depth (OD or D_O) is related to the real depth (RD or D_R) of the stone by its refraction index (Fig. 5).

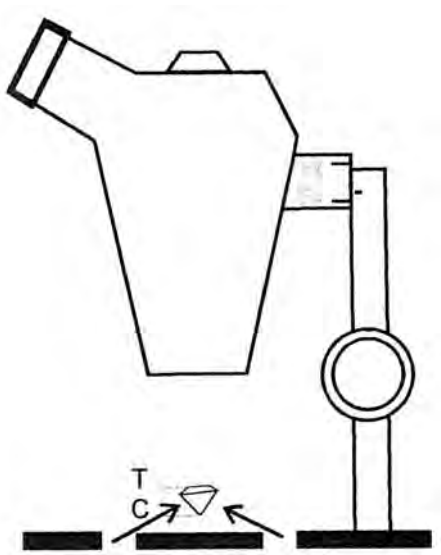


Figure 5. Diagram showing schematically how the 'Duc de Chaulnes' technique is applied to obtain the refractive index or indices of a gem by means of the gemmological microscope. At the highest magnification (i.e., at the smallest depth of focus) the focal point of the culet (F_c) of the gem is recorded by means of a vernier or other mechanical measuring device. Next, the focal point of the table (F_t) is recorded in the same way. The distance between F_c and F_t is the optical or apparent depth (D_O) of the stone. Dividing the real depth (D_R) of the gem by this optical depth (D_O) yields an approximation of the refractive index (RI) of the material.

Using the Duke of Chaulnes method

$$\frac{D_R}{D_O} \approx \text{RI} \qquad (1)$$

This technique is especially interesting when the RI of the stone exceeds the upper limits of the regular Abbé-type refractometer. This latter type needs a contact liquid, which is the limiting factor. Determination of the RI(s) by the Duke of Chaulnes method has no upper limit since no contact exists. In birefringent gems, the accuracy of the measurement

[6] Schneider of Idar-Oberstein has therefore developed a nontoxic vegetal immersion liquid.

can mostly be increased by inserting a polarizer filter.

The Gemmological Microscope as a Dichroscope

Some manufacturers of gemmological microscopes have developed a dichroscopic ocular that has to be inserted just like a regular ocular. Inside this ocular, the field of view is divided into a left and a right field using crossed polarizers. Pleochroic gems viewed through this eyepiece will yield their pleochroic colors.

The Gemmological Microscope as a Spectroscope

Instead of an ocular, a handspectrocope can be inserted into the optical tube of the microscope. The Rayner handspectroscope was specifically built to fit in the eyepieces of most gemmological microscopes. A modern version of this construction consists of a fiber optic light guide placed into the ocular's tube, which is linked to a small PC-based spectrometer. Other more sophisticated applications, such as micro-FTIR spectroanalysis or micro-Raman laser analysis, use a sophisticated scientific research microscope as a base, but are gaining more and more acceptance for the identification of inclusions and/or treatments.

The Gemmological Microscope Used for Crystal Angle Measurements or 'Schmetzer's Microscopic Techniques'

Schmetzer [15] and Kiefert and Schmetzer [16–18] have described two new microscopic techniques for the gemmological horizontal immersion microscope to discriminate between natural and synthetic gems. This is based on the measurement of angles that exist between structural features such as twinning planes and groups of parallel growth planes which run parallel to the crystal faces. Owing to differences in the growth environments, 'habit'-differences also occur (Fig. 6), and are visible as angular differences.

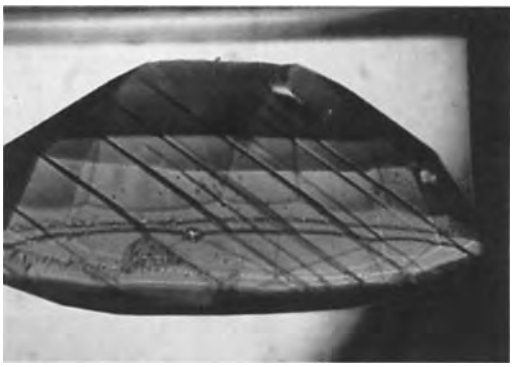


Figure 6. When equipped with polarization filters, crystal angle measurements are possible with the horizontal gemmological microscope in transmitted–immersion illumination. Using the optic axis as the absolute reference, the rotation angle of 90° to the growth planes in this Sri Lankan sapphire identifies them as parallel to *c*(0001). View almost perpendicular to the optic axis, which is oriented diagonally. Crossed polarizers; transmitted–immersion illumination; 20×.

The determination of characteristic structural features such as growth planes or twinning lines comes from the measurement of the angle between the optic axis (OA) (or *c*-axis for uniaxial crystals) and the observed planes. To this effect a special stone holder (Fig. 7) with an angular graduated scale (in degrees) has been developed. The position of the optic axis is first determined in the orthoscopic mode, then the rotation angle to the next crystal plane is read on the graduated scale. A second angular measurement (in-plane angle) can then be taken using an

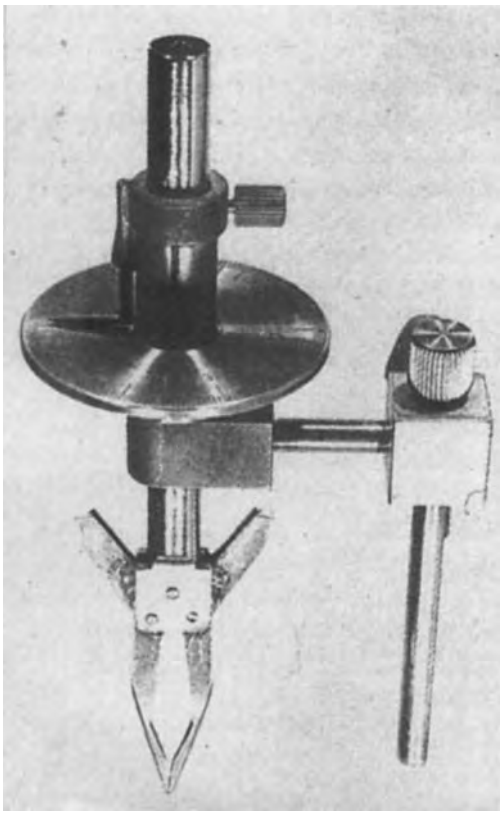


Figure 7. Improved stone holder manufactured by Schneider of Idar-Oberstein with a graduated scale (dial) attached to the vertical axis, allowing easy measurement of the rotation angle between a given plane and the optic axis of the gem.

ocular that has been fitted with a graticule with an angular graduated scale. Measurement of the angles between all visible growth planes allows the identification of these planes and hence the discrimination between different gems from different growth environments.

7.1.2.8 Stone Manipulators

Most manipulators used in gemmological microscopes are of the simple spring-loaded tweezer or stone-holder type. One major disadvantage is that the manipulator only holds the stone. A lot of manipulation is not possible. This can be improved, however, by fitting it onto an $X-Y$ stage which permits precise positioning of the stone during inspection. This manipulation comfort came along in the 1970s with the introduction of the 'Potteratmaster', a sophisticated Swiss manipulator which allowed a round brilliant to spin within three small wheels holding it securely by its girdle. With some accessories, this could also be achieved with fancy shapes. In 1988 the HRD certificates laboratory introduced the 'HRD-gem manipulator', which is, basically, a rotatable vacuum-tweezer. Using silicone-rubber suction pads of different sizes and shapes, almost all sizes and shapes, rough or polished, can be inspected. Using a vacuum, the HRD-gem manipulator can pick up a stone in any required position allowing inspection through the pavilion or the crown-side at will. This results in a very wide field of view. The obvious disadvantage of this manipulator is the necessity to use it linked to a small vacuum pump with an umbilical cord-like vacuum hose.

All previously described stone holders were designed for holding only one stone at a time. For the inspection of whole parcels of small stones these are obviously not very suitable. Van Bockstael [19] mentions the German-made Sortoscope, which solves this problem by using a special sorting tray fitted onto the microscope stand. It includes a sliding white plastic, V-grooved tray which holds the stones and at the same time serves as a diffuse illuminator.

7.1.3 Microscopic Examination of Polished Diamonds

Most of the time, the diamond microscope is used for only one purpose: to grade the quality of polished diamonds. This quality is characterized by the so-called 'four Cs': carat, clarity, color, and cut. The microscope is obviously of no great help when it comes to assessing the carat weight or the color of a diamond. Electronic scales and color reference samples are used for this purpose. The real advantages of the microscope are most obvious when it comes to clarity grading and also, but to a lesser extent, to cut grading.

7.1.3.1 Clarity Grading

The clarity grade of a diamond gives a measure of the number and the importance of the impurities that can be observed in the stone. A basic definition of the different grades was given by the IDC [20] (see Table 3). The internal characteristics are graded as described by the given definitions, which are valid for use by an experienced grader.

At present, four important diamond grading systems are in use (Table 4):

1. The IDC (International Diamond Council) system, which was adopted by the international professional diamond organizations WFDB and IDMA in 1978 and is currently applied by the HRD Certificates Department and three other major grading laboratories.
2. The GIA (Gemmological Institute of America) system applied by the GIA Gem Trade Labs in the U.S.A. and by a number of private gemmological labs throughout the world.
3. The CIBJO (Confédération Internationale de la Bijouterie, Joallerie et Orfèvrerie) system applied by the licensed labs of this international jewelers' organization.
4. The ScanDN (Scandinavian Diamond Nomenclature) system. There are no

Table 3. Clarity grading of polished diamonds[a]

Grade	Definition
Loupe-clean	A diamond is called 'loupe-clean' if, after examination by an experienced grader with a loupe 10× (corrected for spherical and chromatic aberration), under normal light conditions, it has been found free of internal characteristics.
vvs: vvs1, vvs2	Very very small internal characteristic(s), which can be found from 'very hardly' to 'hardly' with a loupe 10×. The size, position, and number of the internal characteristics determine the distinction between vvs1 and vvs2.
vs: vs1, vs2	Very small internal characteristic(s), which can be found from 'fairly hardly' to 'easily' with a loupe 10×. The size, position, and number of the internal characteristics determine the distinction between vs1 and vs2.
si	Small internal characteristic(s), very easy to find with a loupe 10×.[b]
p1	Internal characteristic(s), which can hardly be found with the naked eye through the crown-side of the diamond.
p2	Large and/or frequent internal characteristic(s), easily visible to the naked eye and which reduce(s) the brilliancy of the diamond slightly.
p3	Large and/or frequent internal characteristic(s), very easily seen with the naked eye and which reduce(s) the brilliancy of the diamond.

[a] IDC [20]; [b] since 1991, this IDC grade has been split up into si1 and si2.

Table 4. Comparison of the IDC (e.g., HRD), CIBJO, ScanDN, and GIA clarity grading systems.

IDC (e.g., HRD)	Final decision using loupe 10×									(p = piqué)
								naked eye		
	Loupe-clean	vvs1	vvs2	vs1	vs2	si1	si2	p1	p2	p3
CIBJO	Final decision using loupe 10×									
								naked eye		
	Loupe-clean	vvs1	vvs2	vs1	vs2	si1	si2	p1	p2	p3
ScanDN	Final decision using loupe 10×									
								naked eye		
	Internally flawless (flawless)	vvs1	vvs2	vs1	vs2	si1	si2	p1	p2	p3
GIA	Magnification 10× (loupe-microscope)									(i = imperfect)
								naked eye		
	Internally flawless (flawless)	vvs1	vvs2	vs1	vs2	si1	si2	i1	i2	i3

At present (1996) there seems to be a commercial tendency on the American diamond market to add an si3 grade on the border line between si2 and i1.

actual labs in operation using this system.

Compared with a handheld lens or loupe, the microscope has the advantage of having more than one possible magnification. The most popular types of diamond microscope feature a zoom lens that allows a whole range of magnifications. Furthermore, most diamond microscopes are of the binocular stereomicroscope type. This is not only less strenuous to work with, but it also produces a truly three-dimensional image, which makes the identification of characteristics less cumbersome. Finally, most diamond microscopes feature dark field illumination, which makes even small pinpoints stand out brightly on a dark background. For clarity grading, only 10× magnification is used for the actual assessment. Indeed, the IDC rules clearly specify that the decision on the clarity grade should be made with a 10× loupe. The microscope is only used as an aid to find and identify characteristics. An accessory that allows (semi-)quantification of the effect of the presence of an inclusion on the clarity grade was introduced as early as 1974 by the Diamond Grading Laboratories Inc. in London [21]. This accessory consists of a 'Porton particle size analysis' graticule (Fig. 8) placed in one of the oculars of the microscope. A similar graticule (Fig. 9) was developed later by the HRD Certificates Department. This graticule is used mainly as an internal reference and also as a training device. Since not only the size of an inclusion, but also its contrast to the background, its position in the diamond, and the number of images it produces, defines its grade, a whole set of rules is

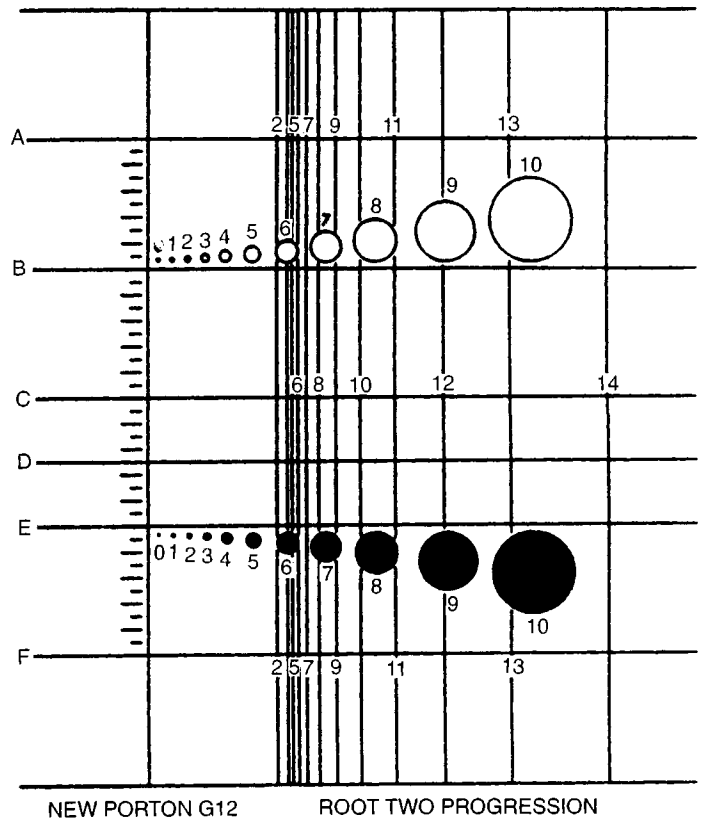


Figure 8. Porton graticule, used to measure the size of inclusions in gemstones.

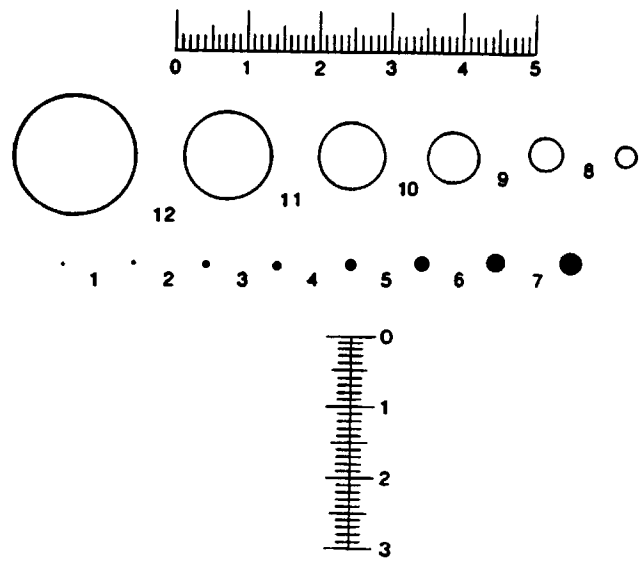
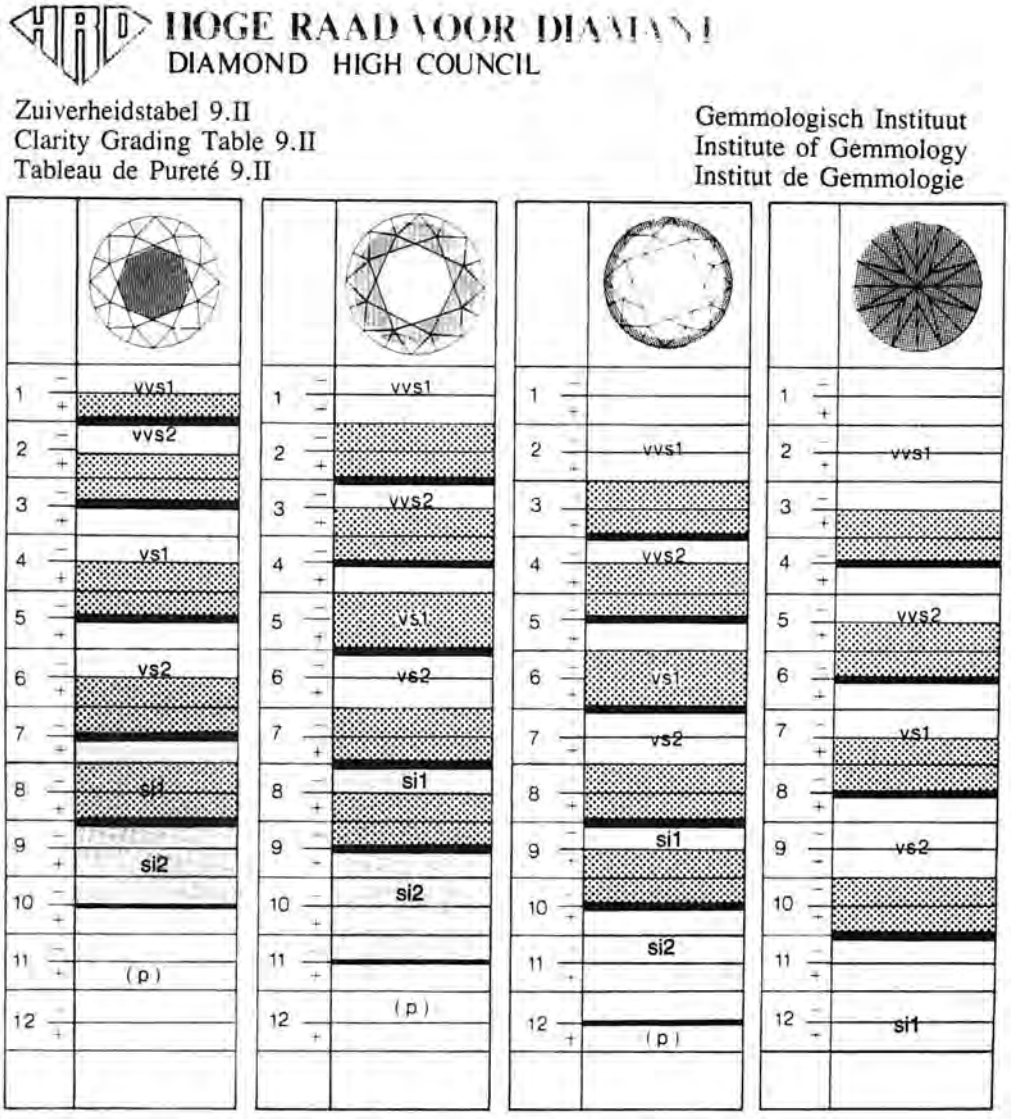


Figure 9. Clarity grading graticule, used by the HRD Certificates Department and by the HRD Institute of Gemmology.

Table 5. HRD clarity grading table.



Contrast	1	2	3	3–4	4	4–5	5
Correction C1	+1	0	−1	−2	−3	−5	−7
Number/code	1	2		3	6		
Correction C2	0	0		+1	+2		

[a] Copyright HRD (1990); [b] this synoptic table translates the four clarity defining parameters (size, contrast to the background, position, and number of reflections) into a clarity grade.

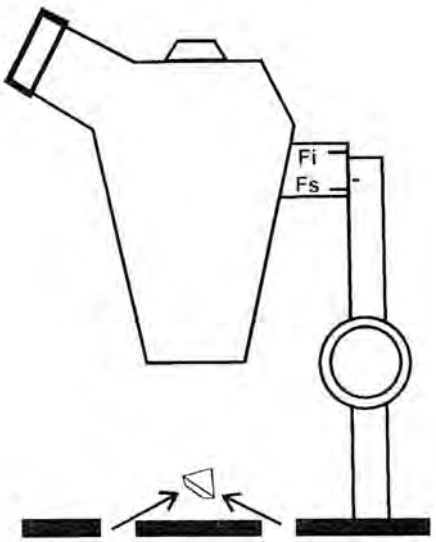


Figure 10. To measure the depth of an inclusion mechanically, the change in the focusing depth ($F_s - F_i$) is recorded with a dial gauge or other mechanical measuring device which is fitted to the microscope (modified from Van Bockstael [22]).

needed to transform these four parameters into a clarity grade. Table 5 relates these parameters to the clarity grade.

7.1.3.2 Measuring the Depth of an Inclusion Using the Diamond Microscope

In order to calculate whether the clarity grade of a diamond can be improved by the removal of an inclusion (repolishing), it is necessary to estimate the distance from the surface of the stone to the inclusion. Most diamond polishers will judge this depth by experience. Depending on the size of the stone and the exact location of the inclusion in the stone, they can balance the expected weight loss and the risk involved in recutting against the expected clarity grade improvement. It is obvious that the observation and estimation of the depth of the inclusion with only a handheld loupe leaves considerable room for error of judgement. For a more accurate observation, the microscope can be used as a superior magnifying instrument. The accuracy of the depth estimation will, however, still depend on the polisher's skill and experience. To overcome this inconvenience, the microscope can be used as a measuring device. This can be done either 'mechanically' or 'optically'. The 'mechanical' depth measurement consists of mechanically measuring the progression of the optical pod between focusing on the surface (focal point F_s) and then on the inclusion (focal point F_i) (see Fig. 10). For this, a dial gauge or other measuring device is fixed onto the microscope stand [22]. The measurement obtained with the dial gauge is only the apparent depth. To obtain the real depth it has to be multiplied by the refractive index ($n = 2.42$ for diamond). The accuracy of the measurement depends largely on the mechanical quality of the focusing thread, and on the optical properties of the microscope lenses. In this respect the depth of focus of the microscope optics at a given magnification is imperative. Since at high magnification the depth of focus is smallest, it is clear that the measurement must be taken at the highest magnification. However, for a given optics the accuracy of the measurement will still be no better than the depth of focus at the highest magnification.

The 'optical' depth measurement avoids both the mechanical error associated with the progression of the optical pod and the optical error produced by the depth of focus of the microscope's optics, which are the limiting factors of the accuracy of the mechanical depth measurement. To arrive at this higher accuracy, the depth of an inclusion is measured in reflection

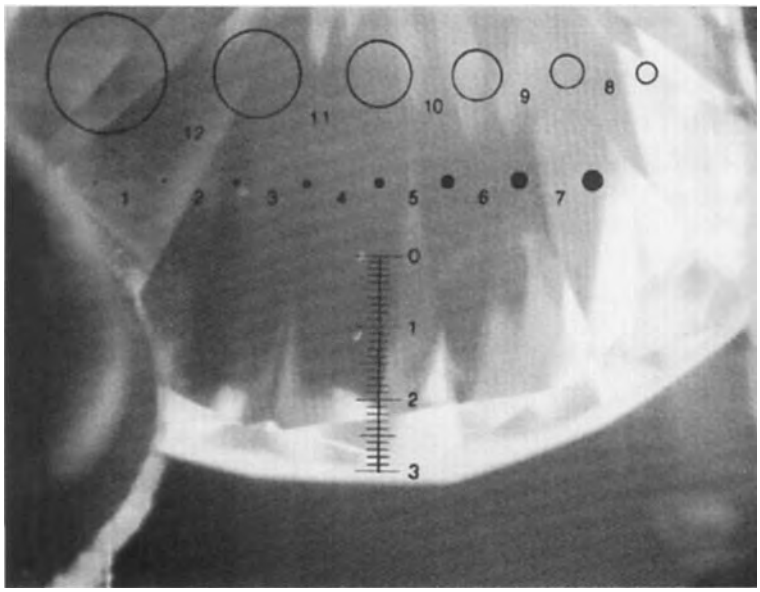


Figure 11. To measure the depth of an inclusion optically, a graticule displaying a linear graduated scale is used to assess the distance between the real image and the reflection of the inclusion. The apparent depth of the inclusion is half this distance.

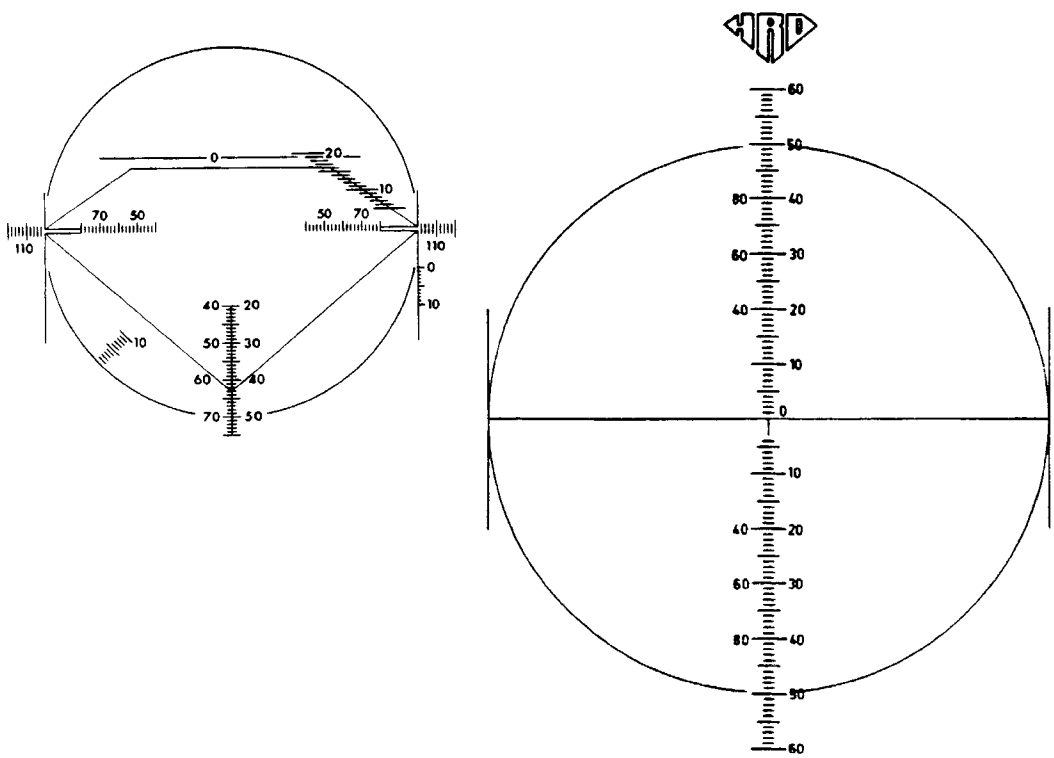


Figure 12. The HRD proportion graticule. The inset (top left) shows the screen used in the standard proportionscope.

with a graticule fitted into the eyepiece of the microscope. The apparent depth of the inclusion is equal to half the distance between the real image of the inclusion and its reflection (see Fig. 11). Multiplication of this figure with a trigonometric constant, a corrective factor for the multiplication used, and, finally, the refractive index of diamond, produces an accurate approximation of the depth of the inclusion.

7.1.3.3 Proportions Grading

For the assessment of the cut, the HRD Certificates Department has developed a 'proportion graticule' (Fig. 12). This allows the grader to measure the crown height, girdle thickness and pavilion depth for a diamond polished in the round brilliant shape. The actual measurement is illustrated in Fig. 13.

The accuracy of the proportions measurement is comparable to, or even better than, that obtained by a regular proportionscope. This is especially so when it is used together with vacuum tweezers, which allow a complete and undisturbed view of the stone. Apart from the proportions, the finish or symmetry deviations can also be quantified using this graticule. The measured results are optimal when the

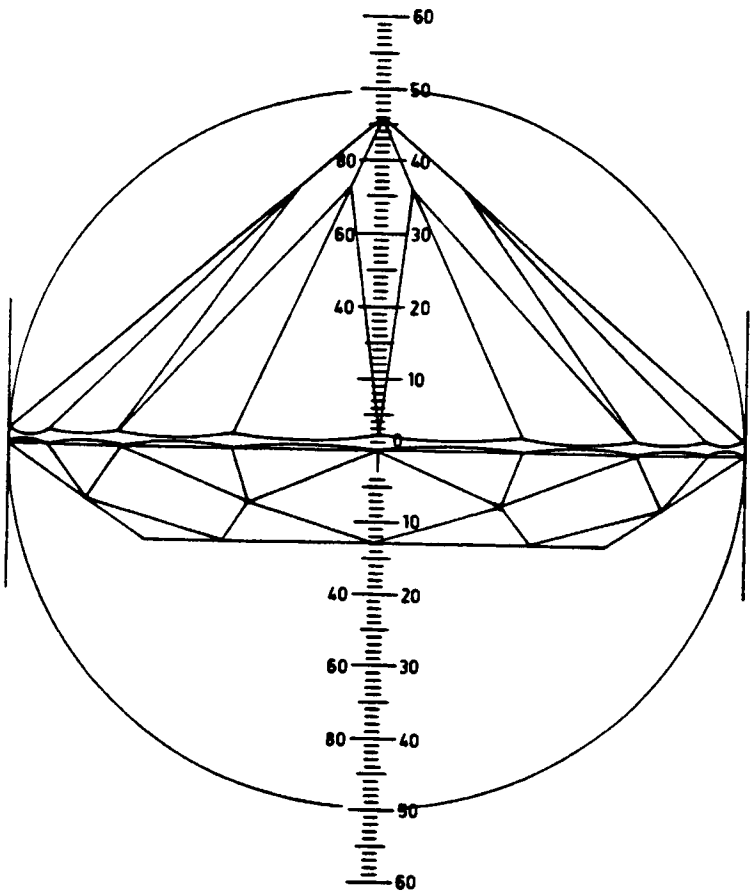


Figure 13. In this example, the crown height of a brilliant is measured using the HRD proportion graticule. At the same time, the girdle thickness can be read from the graduated scale. The crown height is 13% of the total diameter, whereas the girdle thickness indicates 2%.

Figure 14. HRD type IV diamond microscope.

Figure 15. GIA 'Gemolite' gemmological microscope.

overhead daylight lamp of the microscope is used. Two popular diamond microscopes are illustrated in Figs. 14 and 15. In Figs. 16–19, some of the most typical characteristics in diamond are shown, as seen through the microscope.

7.2 The Use of Electron Microscopic Techniques for the Characterization of Gemstones

It is the purpose of this part of the chapter to illustrate with a few examples the use of electron microscopy techniques for the analysis and characterization of gemstones. In most cases, depending on the technique applied, electron microscopy is a rather destructive technique and is therefore not suited for routine application by nonscientists on polished gemstones. Not the observation techniques themselves, but the required specimen preparation techniques are detrimental.

However, the application of electron beam techniques can teach us a lot about the structural characteristics of the material from which the gemstone is tailored and can thus be used for a thorough and unambiguous identification, e.g., for a set of stones of the same provenance. This section consists of two main parts: in the first, introductory part the experimental techniques of electron microscopy, electron diffraction, scanning electron microscopy, and electron beam based microanalysis techniques are briefly outlined and reference is made to the appropriate detailed chapters of Volume 1 of this Handbook; in the second part a few

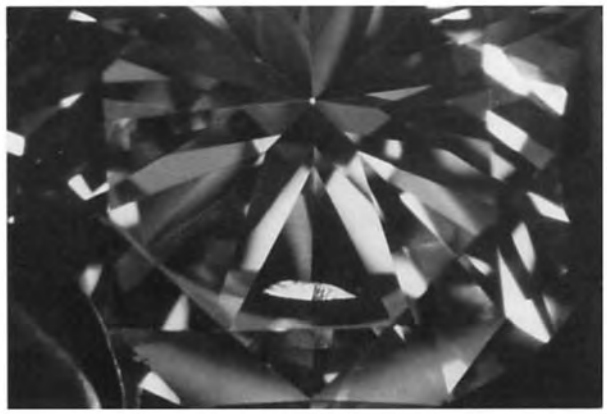


Figure 16. Typical examples of fractures (also called 'gletzes', 'cleavages', or 'feathers') in a diamond.

Figure 17. Typical examples of crystal inclusions in a diamond.

Figure 18. Typical example of a 'cloud' in a diamond. Clouds consist of numerous, very small, pinpoint inclusions.

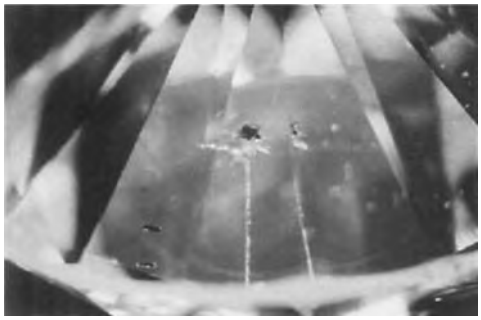


Figure 19. Laser drilling in a diamond. This artificial characteristic was made in order to chemically bleach a very dark inclusion in the stone.

examples are illustrated, mainly from studies performed at the EMAT laboratory in Antwerp (University of Antwerp, RUCA) and from a few other sources from the international scientific literature. Space forbids us to be exhaustive, however.

As pointed out in many of the previous chapters in this Handbook, electron microscopy could be considered chronologically as the successor to optical microscopy, enabling higher magnifications and more detail to be discerned. However, the technique offers more than just a higher magnification, and yields a wealth of structural and chemical information, as we shall see hereafter.

The working principle is easily compared with the optical microscope, one of the main and most sophisticated tools for the nondestructive analysis of gemstones, as was illustrated extensively in Sec. 7.1 of this Chapter. Most important and essentially different is the interaction of the radiation, i.e., light or electrons, with the object. It is this interaction that is responsible for the type of images and contrast that are obtained, and, still more important, the kind of information that can be expected from the image and other recording possibilities. Whereas in light microscopy this interaction is mainly reflection or transmission (for thin samples), the interaction of an electron beam with a specimen can be manifold, especially with a crystalline specimen, which all of our gemstone materials (or nearly all) are.

We will restrict ourselves to the main types of interaction, which are: transmission, elastic scattering (i.e., diffraction at the lattice planes), absorption, secondary electrons or backscattered electrons, characteristic energy losses, and characteristic X-rays emitted from the irradiated area. The first three types of interaction are used in transmission electron microscopy, where a suitable imaging theory is required to interpret the images of the structure, i.e. the defects in the structure, such as stacking faults, twin crystals, grain boundaries, inclusions, etc., as elaborated by Amelinckx in Chap. 4.1.1 of Vol. 1 of this Handbook.

These can nowadays be observed at very high magnifications of up to millions of times and interpreted without ambiguity. By suitable focusing of the lenses, the electron beams diffracted at the specimen can be recorded as a diffraction pattern, which yields information on the structure and the crystallinity of the sample. The characteristic X-rays which are emitted upon irradiating a sample with an electron beam can be analyzed by auxiliary equipment attached to the electron microscope.

Specimens for transmission electron microscopy (TEM) should be very thin: less than 200 nm down to 5 nm, depending on the observation technique or the type of information required. This means that the samples will have to be prepared down to this thickness by specialized techniques

such as sawing, grinding, chemical polishing, ion bombardment, or the like, as have been extensively detailed in Chap. 2.3 of this Volume by Barna et al. These techniques are usually rather destructive and are therefore only or mainly suited for fundamental research on gemstone materials.

A type of electron microscope where the specimen requirements are less drastic, which enables much higher magnification than optical microscopes and a much larger depth of field and range of chemical signals available, is the scanning electron microscope (SEM), the principle of which was treated extensively by Cowley in Chap. 4.2.2 of Vol. 2 of this Handbook. In its simplest mode it utilizes the backscattered and/or secondary electrons from a specimen irradiated with a narrow beam of electrons. The specimen is scanned with this fine beam of electrons and the backscattered or secondary electrons are detected; this signal is recorded on a cathode ray tube (TV screen) synchronously while scanning the specimen. A reliable image of the surface of the specimen is thus obtained. Since no magnifying lenses are required, no lens aberrations spoil the image or the depth of field, which thus becomes very large in comparison with that of the optical microscope. Also, an SEM can be equipped with an X-ray or other microanalysis detector, enabling the determination of the composition at the surface or of particular surface heterogeneities such as, e.g., inclusions or artificially added heterogeneities, on a very small scale.

We shall now illustrate the use of these techniques by means of a few examples of electron microscopy results on gemstone materials.

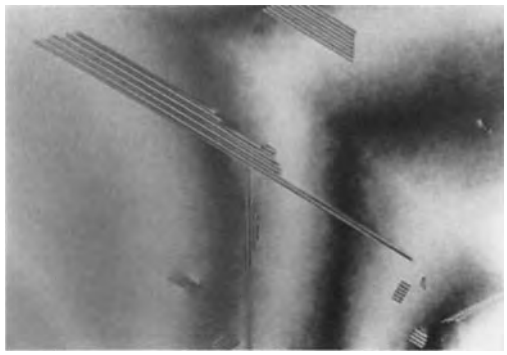


Figure 20. Nitride platelets in diamond observed in diffraction contrast mode (TEM) as a pattern of parallel fringes.

7.2.1 Diamond

An illustration of the use of a scanning electron microscope for a nondestructive study of very small diamonds is shown in Fig. 20, where an approximately 100 μm small, synthetic diamond with an octahedral growth habit is observed at a magnification of 500×. On this figure the very large depth of field of the SEM technique is clearly evident.

Figure 21 shows an example of the combined use of scanning electron microscopy and X-ray microanalytical

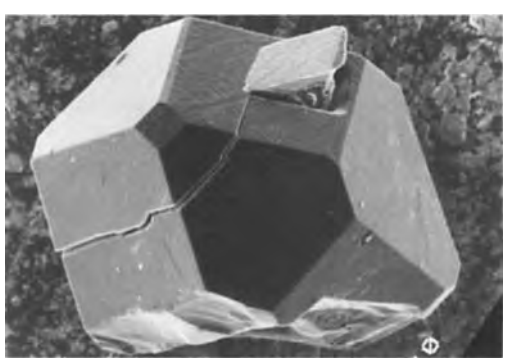


Figure 21. Scanning electron micrograph of a small synthetic diamond (±100 mm). Notice the large depth of field in this SEM image.

techniques, and reveals ion-etched micro-engravings performed on a diamond surface by a procedure developed in Antwerp (Stoneprint). Scanning electron microscopy could also be used to prove the negative relief obtained by ion etching, as shown in Fig. 22a (letter A). The nature of a small accidental particle could be determined by X-ray microanalysis in an SEM to be copper, as evidenced by the spectrum which is illustrated in Fig. 22b.

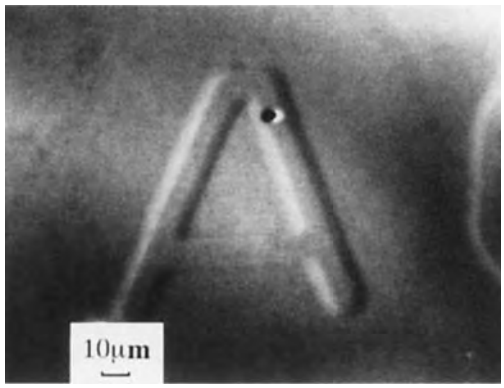
Fundamental studies on the structural defects in natural diamond have revealed the existence of several types of defect often related to the type of diamond: dislocations, dislocation loops, platelets, and voidites [24, 25]. The presence of platelets in diamond can very clearly be observed at high magnification in transmission electron microscopy diffraction contrast (see Chap. 4.1.1 of Vol. 1 of this Handbook by Amelinckx). A thin diamond foil prepared by ion beam thinning intermittent with oxidation down to a thickness of 200 nm clearly reveals these platelets. The fringe patterns are diffraction contrast effects revealing the presence of very fine platelets (Fig. 22). It can be seen that the platelets are lying in strict crystallographic directions.

A detailed electron microscopic study on these platelets was reported by Barry et al. [26]. Figure 23 shows an example of a high resolution image of a platelet. This type of image enabled the authors, by using a number of different techniques such as image matching of perfect diamond images and matching the width and the appearance of the central fringe at the platelet structure image, to conclude that the platelet structure is consistent with a double nitrogen layer in zigzag configuration.

Luyten and co-workers [27, 28] recently reported extensive studies on voidites in type IaB diamonds, their nature, and their presumed relation with transforming platelets and dislocations. It was shown in a thorough analysis using diffraction contrast, electron diffraction, high resolution imaging, and electron energy loss spectroscopy that the voidites are not necessarily related to transforming platelets, but that in both cases the voidite material consists of a nitrogen-containing phase. More details and images of these octahedral-shaped voidites limited by (1 1 1) planes are presented by Bernaerts and Amelinckx in Chap. 1.11 of this Volume. It should also be mentioned that in particular for diamond materials cathodoluminescence originating from the electron beam excitations, as described in Chap. 4, Sec. 1.4.3 of Vol. 1 of this Handbook, by Joy, can yield interesting information on the type of diamond, the defect and (or) impurity center distribution, and their characterization, especially for defects and precipitates that are only present with a low density and therefore sometimes difficult to discern by the above-mentioned contrast techniques.

7.2.2 Quartz

Quartz is the basic material for a number of gemstones (amethyst, citrine, praseolite, etc.). It is well known that this material undergoes a phase transition at 573°C from low (α)-quartz to high (β)-quartz. The existence of a particular type of twins (Dauphiné twins) in quartz is well known but their relation with the nature of the phase transition was only precluded [29].



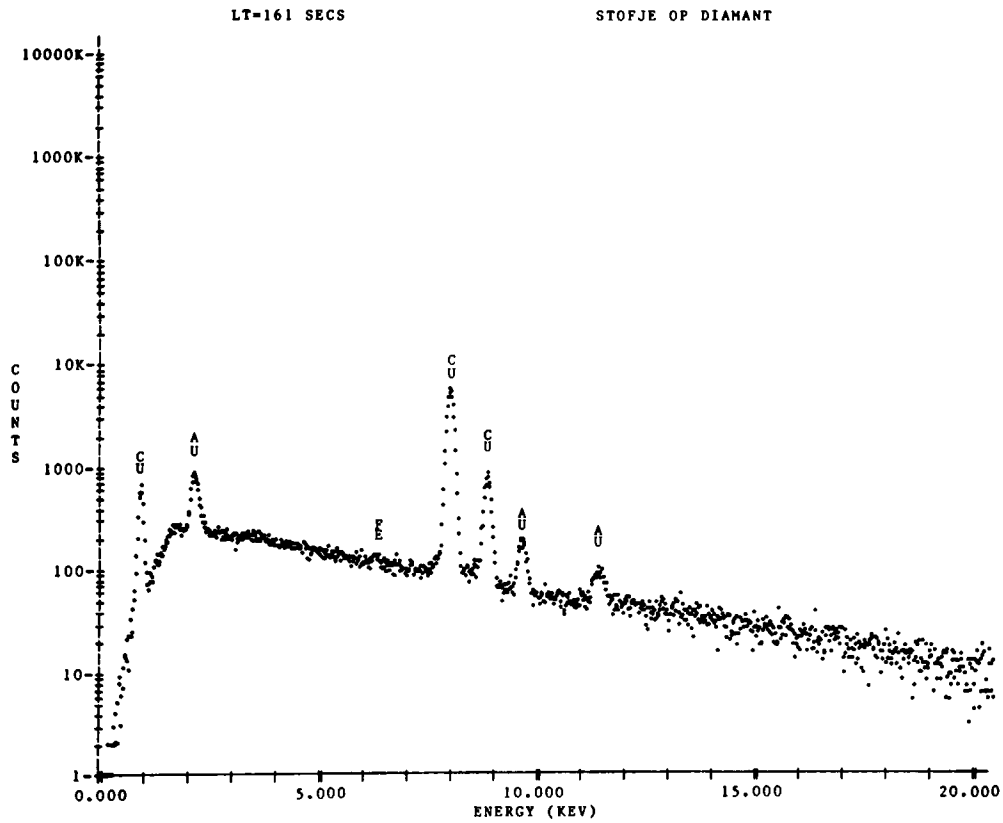


Figure 22. (a) Negative relief lettering on a diamond; the shallow depth relief could be detected by SEM. A small particle of accidental origin was analyzed by X-ray microanalysis and found to be copper from the holder material. (b) The line diagram represents the microanalysis spectrum where, apart from the gold from the intentionally evaporated film, only copper peaks are observed.

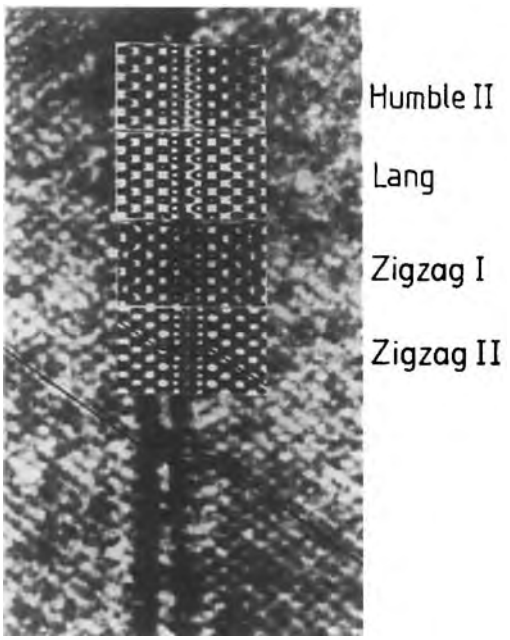


Figure 23. HREM of platelet in diamond; calculated images for various structure models are compared with the experimental image (courtesy of Hutchison).

Figure 24. The phase front between the α- and β-phases in quartz is observed to consist of a regular array of pyramidal Dauphiné twins forming an incommensurate modulated structure, the period of which reduces as the β-phase region is approached. The specimen is close to 573°C; there is a gradient of a few degrees across the area of the photograph.

It has been possible by transmission electron microscopy to characterize the nature of this phase transition as follows [29, 30]. Figure 25 illustrates a time shot of the microstructure taken at the phase transition in quartz. The thinned crystal is heated in a special furnace holder of the electron microscope to a temperature close to the transition temperature of 573°C. Slight electron beam heating is then used to produce a temperature gradient across the observed area shown in Fig. 24.

From the observations and their analysis it could be concluded that at the transition a high density of defects is formed following a regular network with a mesh size that becomes smaller until unresolvable in the β-phase (right part of Fig. 24). These defects could be identified as the well-known Dauphiné twins, which relate two crystallographically equivalent orientations α_1 and α_2 of the α-phase. The observations have led to a novel interpretation of the $\alpha \rightarrow \beta$ transition in quartz, whereby the β-phase can be considered as a time average of the α_1 and α_2 orientations of the α-phase resulting in the existence of an incommensurate phase in between the α and β phases. Theoretical and experimental confirmations and detailed studies have followed this discovery by electron microscopy [31–33].

In amethyst quartz a strong interaction was observed between the Brazilian twins (typical for amethyst) and the Dauphiné

Figure 25. Interaction between Dauphiné twins and Brazilian twins in amethyst quartz observed near the transition temperature.

twins occurring at the phase transition, as shown in the dark field image in Fig. 25 [34].

Various aspects of the nature of the asterism in star quartz are illustrated in Fig. 26. It was shown by Woensdregt et al. [35] that the fine needles causing the asterism are sillimanite and not rutile needles, as generally thought before.

7.2.3 Precious opal

A typical example of the use of scanning electron microscopy is shown in Fig. 27 where a synthetic opal is analyzed. The

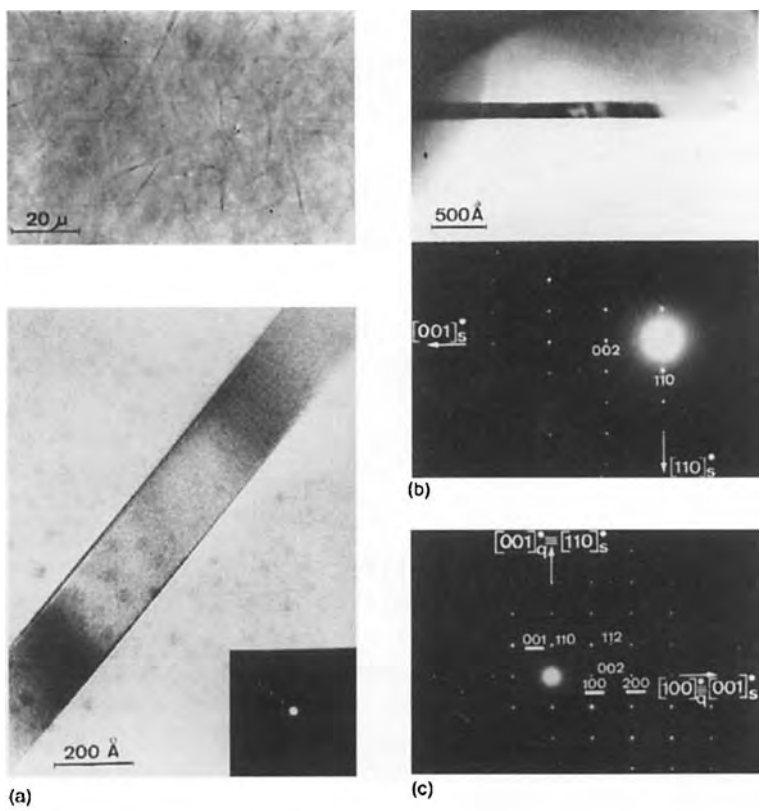


Figure 26. The origin of asterism in star quartz. (a) Transmission electron microscope micrograph of a single sillimanite needle and its electron diffraction pattern. Lattice fringes of $d_{[1\bar{1}0]}$ are visible and the quartz has become amorphous owing to electron beam damage pattern. (b) $[1\bar{1}0]$ electron diffraction pattern of sillimanite. (c) Electron diffraction pattern of the oriented intergrowth of sillimanite (s) and quartz (q). Underlined indices are quartz. [001]* of sillimanite is parallel to [100]* of quartz (courtesy of C. Woensdregt).

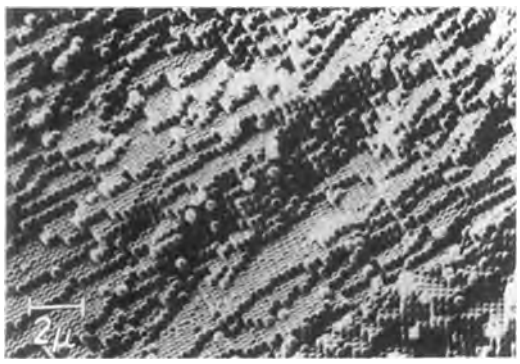


Figure 27. Scanning electron micrograph of a synthetic opal. The regular array of hydrated silica spheres is very evident in this image. The large depth of field results in image sharpness at different heights in the sample.

typical structure of opal as a stacking of small (0.3-mm diameter) spheres in a close-packed arrangement is clearly observed in this SEM image. The advantage of the large depth of field is very evident from this image made by Falter and Liebertz [36]. Object parts several micrometers apart in depth are all in focus. Hexagonal (1 1 1) plane arrays as well as square (1 0 0) plane arrays can be distinguished.

7.2.4 Garnet

As a final example, Fig. 28 shows high resolution images of a garnet crystal thinned by crushing and dispersing crystal fragments on a thin carbon film. A crystal fragment is observed in the high resolution mode, where structural detail of the order of 0.2 nm is observed to confirm the perfect structure.

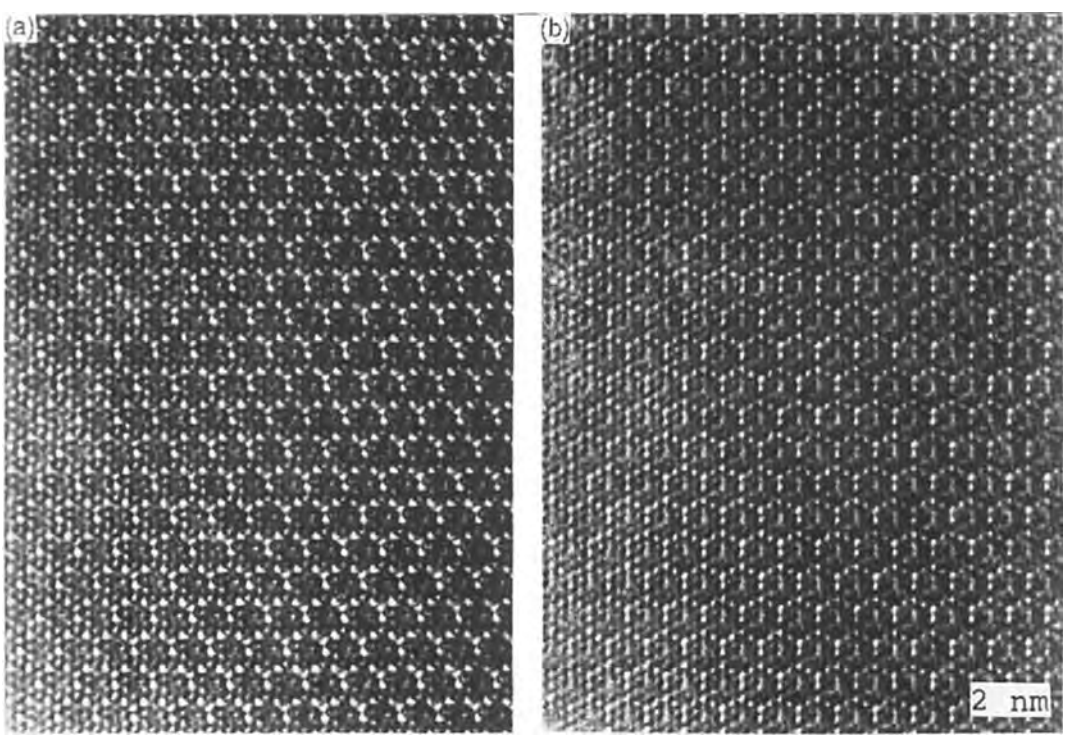


Figure 28. Structural images of a garnet crystal.

7.3 Conclusions

In view of the complexity of this subject, the authors are aware of the fragmentary character of this chapter. They hope, however, that potentialities and disadvantages of the utilized types of microscopy and related techniques can be understood and are well illustrated by the different examples of electron microscopy studies of gemstones that are presented. Semiprecious stones and minerals, being related in that both are naturally grown materials, are often very hard, and contain similar types of growth defects, can mostly be studied by the same methodology. The reader is therefore also referred to the extensive chapter by Barber on this subject (Chap. 1.2 of this Volume).

7.4 References

[1] R. Webster, *Gems: Their Sources, Descriptions and Identification*, 4th ed. (revised by B. Anderson), Butterworths, London **1983**.
[2] J. E. Arem, *Color Encyclopedia of Gemstones*, Van Nostrand Reinhold, New York **1977**.
[3] Anonymous, *Gems Gemology* **1934**, 127.
[4] E. J. Gübelin, *Internal World of Gemstones*, 3rd ed., ABC Edition, Zurich **1983**.
[5] E. J. Gübelin, J. I. Koivula, *Photoatlas of Inclusions in Gemstones*, 2nd ed., ABC Edition, Zurich **1992**.
[6] J. I. Koivula, *Gems Gemology* **1986**, *22*, 152.
[7] P. W. Van Maaren, *Statische en kinematische microscopie*, Elsevier, Amsterdam **1976**.
[8] J. I. Koivula, *Gems Gemology* **1984**, *20*, 101.
[9] J. Figueras, *Boletin Assoc. Espanol Gemologia* **1995**, *36*, 8.
[10] J. I. Koivula, *Gems Gemology* **1982**, *18*, 160.
[11] E. J. Gübelin, *J. Gemmology* **1957**, *6*, 151.
[12] I. J. Koivula, *Gems Gemology* **1982**, *18*, 83.
[13] N. I. Sax, *Dangerous Properties of Industrial Materials*, 5th ed., Van Nostrand Reinhold, New York **1979**.
[14] G. Lenzen, *Edelsteinbestimmung mit gemmologischen Geräten*, Verlagsbuchhandlung Elisabeth Lenzen, Kirschweiler **1984**.
[15] K. Schmetzer, *Natürliche und synthetische Rubine*, E. Schweizerbart'sche Verlagsbuchhandlung (Nägele u. Obermiller), Stuttgart **1986**.
[16] L. Kiefert, K. Schmetzer, *J. Gemmology* **1991**, *22*, 344.
[17] L. Kiefert, K. Schmetzer, *J. Gemmology* **1991**, *22*, 427.
[18] L. Kiefert, K. Schmetzer, *J. Gemmology* **1991**, *22*, 471.
[19] M. Van Bockstael, *Antwerp Gems* **1990**, *1*, 39.
[20] IDC, *IDC-Rules for Grading Polished Diamonds*, Internal Report, International Diamond Council, HRD Antwerp **1979**.
[21] K. F. Burr, *D.G.L. Technical Report*, D.G.L. Laboratories, London **1974**, *1*.
[22] M. Van Bockstael, *Antwerp Gems* **1992**, *3*, 18.
[23] R. Crowningshield, C. Hurlbut, C. W. Fryer, *Gems Gemology* **1986**, *22*, 130.
[24] T. Evans, C. Phaal, *Proc. Roy. Soc. London* **1962**, *A270*, 538.
[25] N. Sumida, A. R. Lang, *Proc. Roy. Soc. London* **1988**, *A419*, 235.
[26] J. C. Barry, L. A. Bursill, J. L. Hutchison, *Phil. Mag.* **1985**, *A51*, 15.
[27] W. Luyten, G. Van Tendeloo, P. J. Fullen, *Phil. Mag.* **1994**, *A69*, 767.
[28] T. Evans, I. Kiwlawi, W. Luyten, G. Van Tendeloo, G. S. Woods, *Proc. Roy. Soc. London* **1995**, *A449*, 295.
[29] R. A. Young, Defense Dept. Doc. Cent. Rep. No. 1D-196.235 **1962**.
[30] G. Van Tendeloo, J. Van Landuyt, S. Amelinckx, *Phys. Status Solidi (a)* **1975**, *30*, K11.
[31] J. Van Landuyt, G. Van Tendeloo, S. Amelinckx, M. Walker, *Phys. Rev.* **1995**, *B31*, 2986.
[32] T. A. Aslanyan, A. D. Levanyuk, *Sov. Phys. JETP Lett.* **1978**, *28*, 70.
[33] G. Dolino, J. P. Bachheimer, B. Berge, C. M. E. Zeyen, G. Van Tendeloo, J. Van Landuyt, S. Amelinckx, *J. Phys.* **1984**, *45*, 901.
[34] L. Van Goethem, J. Van Landuyt, S. Amelinckx, *Phys. Status Solidi (a)* **1977**, *41*, 129.
[35] C. F. Woensdregt, W. Weibel, R. Wessicken, *Schweiz. Miner. Petrogr. Mitt.* **1980**, *60*, 129.
[36] M. Falter, J. Liebertz, *Z. Dtsch. Gemm. Gesellschaft* **1978**, *27*, 134.

Handbook of Microscopy
S. Amelinckx, D. van Dyck,J. van Landuyt , G. van Tendeloo

8 Superconducting Ceramics

8.1 Introduction

Since the discovery by Bednorz and Müller of superconductivity around 30 K in the barium–lanthanum–copper–oxygen system [1], a large number—several hundred—of superconducting compounds with transition temperatures as high as 133 K [2] have been synthesized. Many of these compounds would never have been discovered without the help of various microscopy techniques. In particular, electron microscopy and techniques related to electron microscopy such as energy dispersive X-ray (EDX) spectroscopy or electron loss spectroscopy (EELS) were essential in the evolution of superconductivity research.

Optical microscopy and scanning electron microscopy have strongly contributed to the obtention of single phases of the superconducting compounds and the determination of the morphology as a function of the growth conditions. Polarized light microscopy is one of the few techniques to directly observe the magnetic flux or flux motion in thin superconducting films. Different electron microscopy techniques have played a crucial role in the synthesis and the optimization of the superconducting phases. Sometimes superconducting minority phases—recall the 'green' phase and the 'black' phase in yttrium–barium–copper–oxygen (YBCO) [3]—have been detected and their structure determined with the help of electron microscopy. In many cases, as we will see further, synthesis of new materials and characterization by electron microscopy go hand in hand. Atomic force microscopy (AFM) and scanning tunneling microscopy (STM) are highly important in the elaboration of detail in the surface of bulk superconductors as well as of thin films in different stages of production. Other techniques, such as laser scanning microscopy, positron anihilation, micro-Raman imaging, field electron emission microscopy, and scanning Auger microscopy, are less used techniques, but which certainly provide useful information; their specific use will be highlighted.

In this contribution we will try to illustrate the power of the different imaging and microscopy techniques to solve particular problems in the field of superconductivity. However, it is not the aim to give a survey of all the scientific contributions based on the different imaging techniques. *Physica C*, a journal entirely devoted to superconductivity and superconducting materials, lists over 500 papers which over the past 5 years have been entirely or partially devoted to some

form of microscopy. This shows the impossibility of making a complete survey, but on the other hand also the strength and the importance of direct imaging in superconductivity research.

We will first consider the strength of the different imaging techniques in the field of superconductivity research and then consider a number of problems (physical, chemical and engineering) and see how they can be solved or solutions suggested by the different techniques.

8.2 Imaging Superconducting Ceramics

8.2.1 Morphology Studies by Optical Microscopy

Optical microscopy (OM) is an easy and fast technique to study the morphology of newly synthesized material. Particularly when special requirements have to be met concerning the alignment of grains, or texturing, OM is the technique of choice. The dimensions of most grains is such that the resolution of OM is not the limiting factor. In contrast to other techniques treated in this section, the term 'optical microscopy' seldom appears in the title of a superconductivity paper, indicating that OM is considered a supporting technique in the processing or texturing superconducting materials.

Highly oriented YBCO bulk material can be produced by a melt process or melt texturing ([4–8]); the development and final morphology was followed in all cases by OM. Similarly, single crystals with high critical current values can be grown, OM evaluating the dispersion of foreign grains [9]. The microstructure of different oxides (superconducting or not) in the YBCO phase diagram and the formation conditions of the different phases was monitored by optical inspection of the ceramic samples [10]. The effects of BaSnO additions to melt powder, melt grown (MPMG)-processed YBCO, in an effort to increase the pinning properties of melt textured YBCO, is characterized by a combination of OM, scanning electron microscopy (SEM), X-ray diffraction, and EDX [11]; as usual, a combination of different techniques provides the most complete information.

The crystal growth and decomposition of bismuth-based oxide superconductors $Bi_2Sr_2Ca_nCu_{n+1}O_z$ (BSCCO) has been followed by Komatsu et al. [12] by high-temperature OM. They studied the growth process by slowly cooling a liquid drop on an MgO substrate. The formation of the high-T_c phase by a peritectic reaction between liquid and solid is observed. A sequence of micrographs of the crystals that appear between 865 and 845°C is shown in Fig. 1.

8.2.2 Scanning Electron Microscopy, Scanning Tunnelling Microscopy, and Atomic Force Microscopy Surface Studies

SEM is in a sense the prolongation of OM; it is more a supporting technique than a basic research tool. Only the resolution is better than OM, and the focus depth is infinite (see Vol. I of this Handbook). Like OM, SEM is often used to characterize the surface structure of superconducting

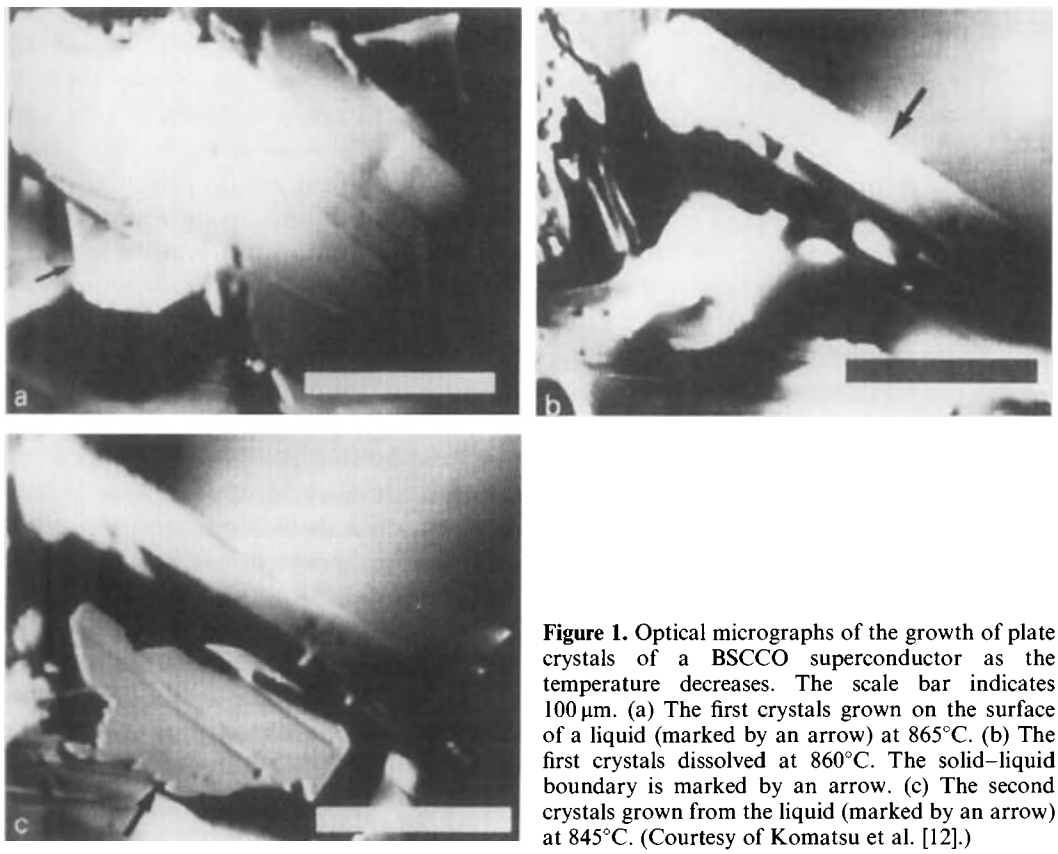


Figure 1. Optical micrographs of the growth of plate crystals of a BSCCO superconductor as the temperature decreases. The scale bar indicates 100 μm. (a) The first crystals grown on the surface of a liquid (marked by an arrow) at 865°C. (b) The first crystals dissolved at 860°C. The solid–liquid boundary is marked by an arrow. (c) The second crystals grown from the liquid (marked by an arrow) at 845°C. (Courtesy of Komatsu et al. [12].)

materials. Unlike OM, however, the samples need not to be polished; because of the enormous depth profile of SEM, it is the preferred technique for the fast inspection of sample composition, grain distribution, crystal size or morphology. The resolution limit—for practical use—is about 10 nm; for most applications this is more than sufficient; when more detailed information is needed there is AFM and STM. The field of applications is therefore more or less the same as in OM, that is, an aid to the synthesis of new superconducting materials and the processing of known superconductors. An advantage of SEM—apart from its depth profile—is the possibility to combine it with energy analysis of X rays, providing compositional information on the area imaged; this has been used, for example, by Su et al. [13] and Lin et al. [14] in studies of single-crystal and sintered wires of YBCO. An interesting study by Lin and Liang [15] investigates the crystallographic defects in YBCO single crystals, revealed by chemical etching with several acids. Different etch pits are observed: some are of the spiral growth type, others have more of a Hopper-like morphology [15]. An example is shown in Fig. 2. For superconducting materials which have synthesis problems, such as $TlSr_2Ca_2Cu_3O_x$ or

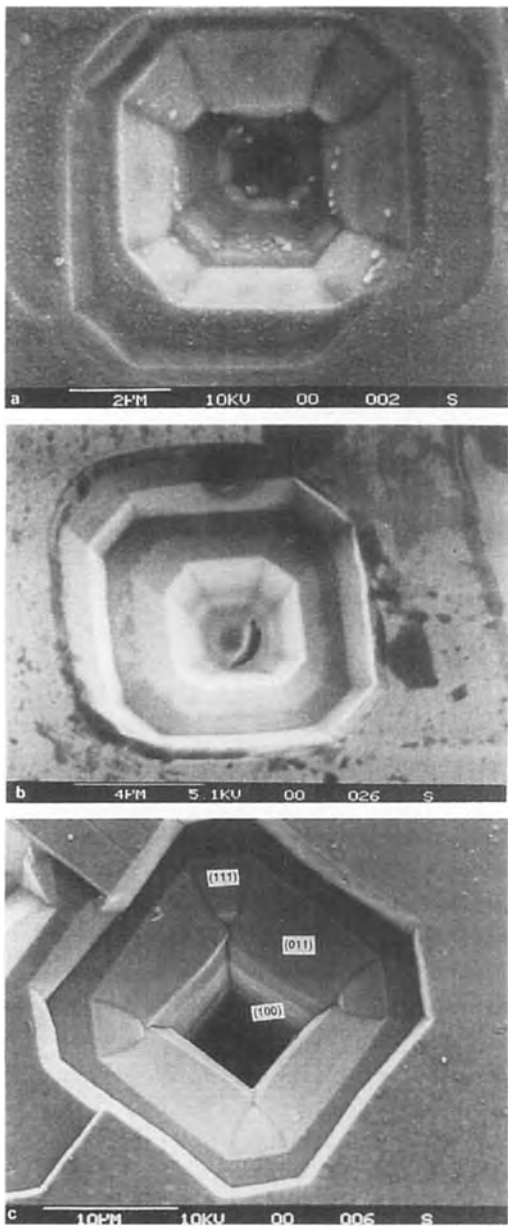


Figure 2. Different morphologies of the typical dislocation etch pits in the (001) plane of $YBa_2CU_3O_{7-x}$ single crystals. (a) By treatment in solution (1 : 4 HNO_3/acetic acid) for 5 min. (b) By treatment in acetic acid for 2 h. (c) By treatment in acetic acid for 4 h. (Courtesy of Lin and Liang [15].)

$Tl_{0.6}V_{0.5}Sr_2Ca_2Cu_3O_x$, SEM helps in establishing the conditions for obtaining single-phase material [16, 17]. As BSCCO is a very anisotropic crystal, it is very promising for thin or thick film deposition. The development of the aligned microstructure during the melt processing of doctor-bladed films can even be followed by SEM [18].

STM or AFM also study the surface, and to some extent overlap with SEM, but they offer a higher resolution; the drawback being that generally a smaller area can be studied. Detailed information on the surface morphology in single crystals of YBCO superconductors has been obtained by, for example, Van de Leemput et al. [19] and by Tanaka et al. [20]. Kroener et al. took advantage of the possibilities of AFM to study the surface and the expansion of YBCO thin films after being patterned by proton irradiation [21]. STM is particularly suited to give information on the growth mechanism of single crystals or thin films. Different reports based on STM have reported on screw dislocation growth [22–24], island growth [25], or layered etching [26]. Apparently there is still some debate on the exact growth mechanism, and some questions remain unsolved: under which conditions do the films develop screw dislocations, and when do they show a layer-by-layer growth behaviour? Zhu et al. [27] have devoted a complete study to this problem for YBCO films as well as for BSCCO films. An image is shown in Fig. 3, where the surface is seen on a unit cell scale and where, under specific conditions, screw dislocations form [27]. Atomic resolution images at room temperature and in an argon atmosphere on BSCCO

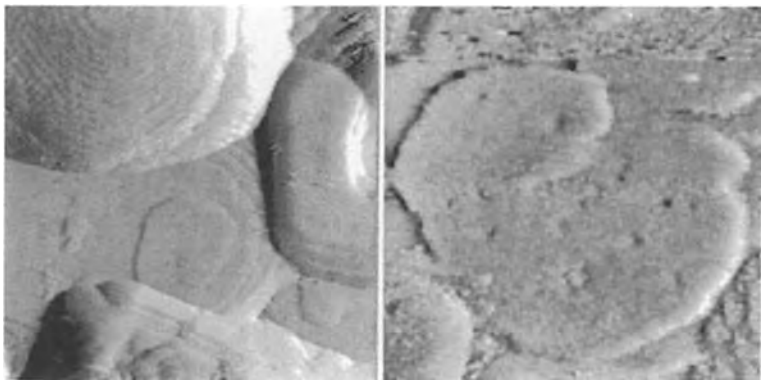


Figure 3. (a) STM image of a $YBa_2Cu_3O_{7-x}$ film grown on a flat substrate; screw dislocations nucleated at the positions of stacking faults or anti-phase domains. The size of the image is 300 × 300 nm. (b) Enlarged image of a single screw dislocation; the size of the image is 120 × 90 nm, the vertical height of the edge is one *c* parameter (1.2 nm). (Courtesy of X. Zhu et al. [27].)

superconductors (2201 as well as 2212) are shown by Boekholt et al. [28] for the [001] cleavage plane. The Bi–Bi (or O–O) distance in the plane is clearly seen. For the 2212 compound, which gives the better resolution because of the better conducting properties, the typical modulation along the *b* axis (see below for the modulated structure) is also revealed in the surface image. Even higher resolution images of the top BiO layer in the cleaved 2212 BSCCO superconductor are reported by Inoue et al. [29], when observed under high-vacuum conditions. The images show the surface modulated structure of the bismuth lattice—already observed by Kirk et al. [30] and Shih et al. [31]—with missing atom rows very clearly (Fig. 4). The modulation at the surface is formed by the ordering of the higher and the lower atoms, and the periodicity is five units, although the bulk modulation is incommensurate (see below). The presence of interstitial oxygen in the BiO layer is observed when 50% of the calcium is replaced by dysprosium [32].

8.2.3 Transmission Electron Microscopy Structural Studies

Since the high discovery of high-T_c superconducting ceramics by Bednorz and Müller [1], and particularly after

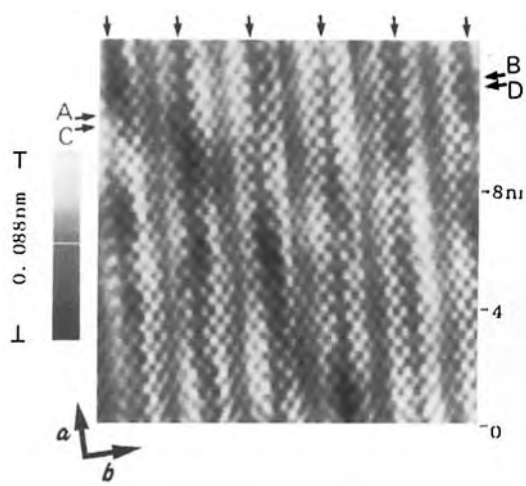


Figure 4. Top STM view (13.2 × 13.2 nm^2) of a cleaved $Bi_2Sr_2CaCu_2O_{8+\delta}$ surface taken at a tip-to-sample bias of −1.02 V and a tunneling current of 0.69 nA. The image was taken in the constant-current mode. The topographic height is shown in grey scale. The image noise is reduced with a low-pass filter. The missing atom rows are indicated by arrows.

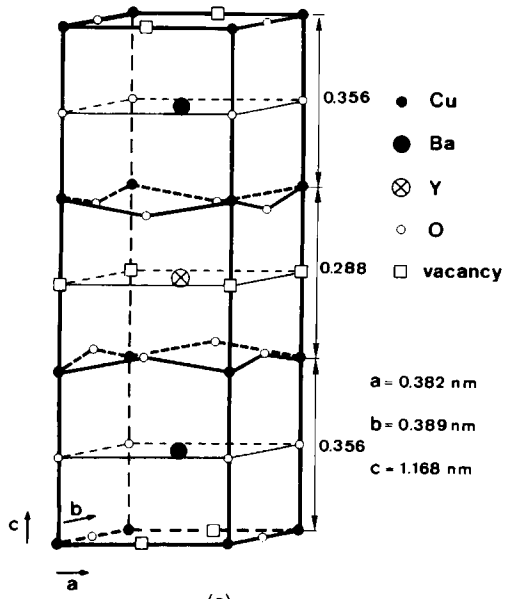
superconductivity was found in $YBa_2Cu_3O_7$ [2], electron microscopy has played a major role in the study of superconducting materials—not only in the structural characterization of the materials, but also in the search for the flux pinning mechanism and the development of new superconducting materials. Electron microscopy has several advantages over other structural techniques, the most important being that it combines information from reciprocal space with local information from direct space. In recent years, real space imaging has improved to the level of atomic imaging (see Vol. II of this Handbook), and electron diffraction patterns can be obtained from nanometer scale areas. This makes transmission electron microscopy (TEM) (where we consider direct imaging together with electron diffraction as a unit) extremely powerful for studying small precipitates or defective structures, which for superconducting ceramics in particular are very important.

The literature on TEM studies of superconducting materials over the last 8 years is so extensive—several thousands of papers—that it is an illusion to aim for completeness. We will consider the $YBa_2Cu_3O_7$-type superconductor (1-2-3 YBCO) in somewhat more detail because it is not only the most studied material but it is also the superconductor which is closest to real applications.

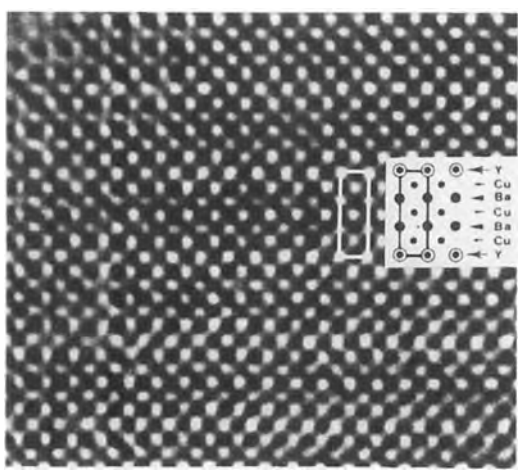
8.2.3.1 Structural Defects in the $YBa_2Cu_3O_7$ Superconductor

(110) Twin Interfaces

The structure of 1-2-3 YBCO is based on the perovskite unit cell with the smaller copper ions occupying the corners of the unit cell and the larger barium or yttrium ions occupying the body-centered position; the oxygen ions occupy edge centers. The unit cell consists of three stacked perovskite unit cells along the c direction, with a regular alternation of two barium and one yttrium ions. In the fully oxygenated state ($YBa_2Cu_3O_{7-x}$) with $x = 0$, the unit cell is orthorhombic with $a = 0.382\,\text{nm}$, $b = 0.388\,\text{nm}$, and $c = 1.168\,\text{nm}$. The small orthorhombicity is due to the fact that there are copper–oxygen chains aligned along the b axis with vacant oxygen sites in between at the $(\frac{1}{2}00)$ positions. The structure is schematically shown in Fig. 5a, and a corresponding high-resolution electron microscopy (HREM) image along [100] is shown in Fig. 5b. The fact that the structure of the fully oxidized state is pseudo-tetragonal, introduces twinning on the coherent (110) or $(1\bar{1}0)$ planes; such twinning will minimize the strain introduced by the phase transformation between the tetragonal high-temperature phase and the orthorhombic low-temperature phase. The detailed structure of twin boundaries in $YBa_2CU_3O_{7-x}$ has been the subject of a large amount of speculation, and several models have been proposed. Some of these models assume the twin boundaries to be perfectly sharp up to the interface [33–35] and propose a deformed oxygen coordination around the copper atoms located in the boundary plane [36, 37]. The heavy ion sublattice is assumed to be continuous across the boundary, apart from a small change in orientation due to the twinning shear. Other models assume the boundaries to be diffuse as a result of the presence of several vacancy-containing layers parallel with the interface [38]. It



(a)



(b)

Figure 5. (a) Schematic representation of the $YBa_2Cu_3O_7$ structure. (b) HREM image along the [100] direction; there is a clear one-to-one relationship between the bright (or black) dot configuration and the heavy ion configuration in (a).

was further suggested that transition metal ions such as iron and cobalt, when substituted for copper, cause fine-scale twinning by introducing an octahedral coordination of oxygen ions around them differing from the linear arrangement in the Cu–O–Cu–O chains [39–44].

In the initial and pioneering work of the EMAT group [45, 46], we have studied 'in situ' their motion during the tetragonal to orthorhombic phase transition. These observations lead to the suggestion that the motion is a combination of a diffusion process and a shear process [47].

An ideally sharp twin boundary would look as shown schematically in Fig. 6a, that is, the orientation of the [100] (or [010]) rows changes abruptly over an angle $F = 90° - 2\tan^{-1} b_0/a_0$. Should the boundary consist of a narrow strip of tetragonal $YBa_2CU_3O_6$ structure, resulting from the presence of a number of oxygen vacancy layers along the boundary, the [100] rows would have the shape represented schematically in Fig. 6b. The orientation now changes in two steps, and a short segment having an intermediate orientation would be formed.

A perfectly sharp boundary is shown in Fig. 7a; such interfaces are observed in well-annealed, nearly stochiometric $YBa_2Cu_3O_7$. Most HREM images published in the literature exhibit a contrast as in Fig. 7b, where the twin interface seems to be spread out over several unit cells, and where the contrast is locally different from the background contrast. There is no evidence for a boundary containing a strip of O_6 structure, such as shown in Fig. 6b; moreover, if such a structure were present we would expect a local change in the interplanar spacing of the (110) lattice planes parallel with the boundary. Applying different digital image-processing methods to high-quality images we have not been able to find evidence for such a twin boundary

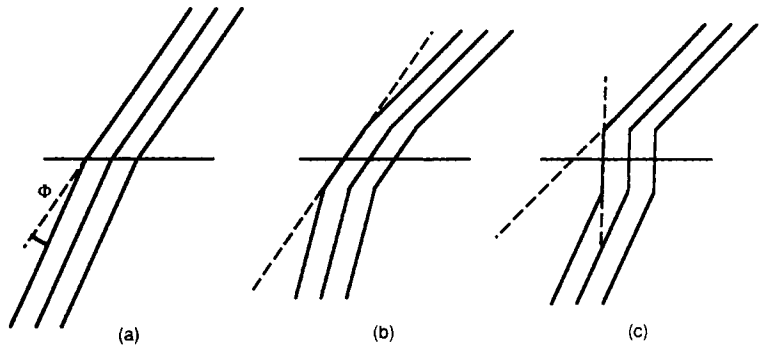


Figure 6. Schematic representation of the orientation change of the [100] lattice rows at a (110) twin boundary of $YBa_2Cu_3O_7$ for two different situations: (a) an ideally sharp boundary and (b) a narrow strip of tetragonal material is assumed to be present along the boundary. (c) A microtwin configuration. The angles are exaggerated to emphasize the characteristic features.

relaxation, putting an upper limit to the spacing change of 2×10^{-3} nm.

In YBCO samples heated 'in situ' inside the electron microscope, the boundary images are in general broadened during the few minutes after the cooling process, especially in samples which have been cycled through the orthorhombic to tetragonal transition and exhibit the short-range ordered ortho II structure due to oxygen loss. These are due, we believe, to intrinsically broad boundaries. Such boundaries sharpen up on further annealing and finally they may become sharp again, provided not too much oxygen is lost. It is reasonable to expect that in such samples the concentration of vacancies is enhanced along the twin boundaries, causing a broadened image.

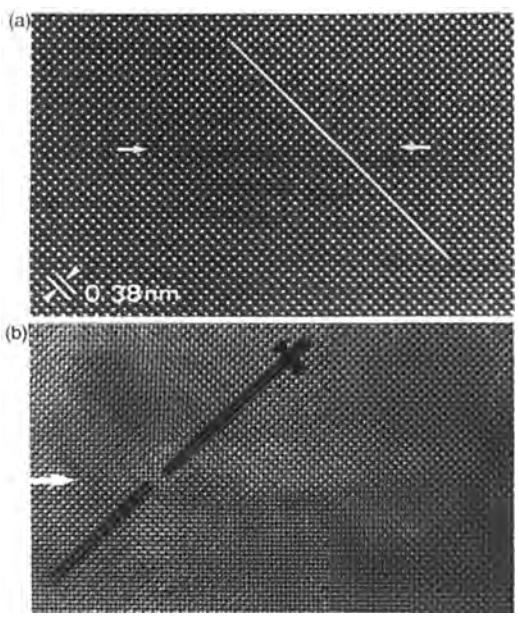


Figure 7. (a) High-resolution image of an 'ideally' sharp (110) twin boundary in $YBa_2Cu_3O_7$. Note the sudden change in orientation of the 100 lattice rows at the line indicated by arrows. (b) HREM image of an oxygen-deficient material, showing a broadened twin boundary in nonstoichiometric $YBa_2Cu_3O_{7-x}$.

The general conclusion is that in well-annealed samples with an oxygen content close to stoichiometric, twin boundaries are sharp on an atomic scale, that is, one or two atomic layers, whereas they are broadened in the oxygen-deficient samples, their width still being a function of the annealing treatment. However, in the interpretation of HREM images of twin boundaries in the literature, one has to be aware of the fact that possible misalignment or the presence of microtwins may broaden the image. The observation of

broadened twin images therefore does not necessarily mean that the boundary is actually broadened. Trivial causes of image broadening have to be eliminated before one can conclude that a boundary is 'intrinsically' broadened.

'In situ' electron microscopy studies of the tetragonal to orthorhombic phase transition has shown that on cooling through the transition, in the initial stage microdomain nuclei in the twin configuration are formed with two families of coherent interfaces, (110) and (1$\bar{1}$0). These give rise to a 'tweed' structure, due to the overlap of narrow microtwins [47], [44]. Such a tweed 'texture' is shown in Fig. 8. The corresponding diffraction pattern then exhibits overall fourfold symmetry; in particular, there is no spot splitting. However, the diffraction spots are not circular but form small crosses extending in the [110] and [1$\bar{1}$0] directions (Fig. 9). It was therefore suggested that the crystal locally still had orthorhombic symmetry whilst macroscopically it shows fourfold symmetry. It was further deduced from video recording that as soon as rather sharp observable twin boundaries are formed, the widening of the twin bands occurs by the lengthwise propagation of ledges along the twin interfaces.

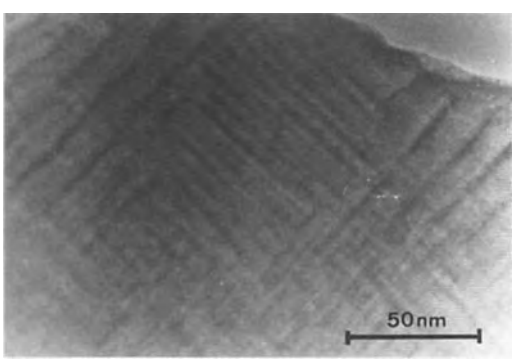


Figure 8. Coarse 'tweed' structure due to interweaved twin boundaries in a beam-heated sample.

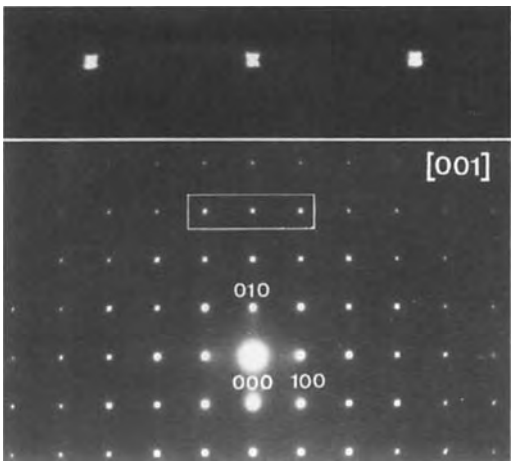


Figure 9. Diffraction pattern produced by a sample exhibiting a 'tweedy' image (see Fig. 8). Note the streaked shape of the diffraction spots in the enlargement.

In the initial stages, that is, during the formation of microdomains, the twin formation proceeds essentially by a rapid diffusion process, aided by the presence of structural vacancies in the CuO_{1-x} layers. In this stage the twin interfaces produce diffuse images, as can be judged from Fig. 8. The widening of the twin bands occurs apparently by a shear mechanism involving the propagation of twinning dislocations, driven by the ordering stresses. Since $b_0 > a_T$ and $a_0 < a_T$ ordering causes an expansion in the b_0 direction and a contraction in the a_0 direction. The cooperative deformation resulting from an ordered domain thus causes a shear stress, as represented schematically in Fig. 10c, which acts on the twinning dislocations and induces their conservative motion along the interfaces. However, in the movement of a twin plane a pure twinning shear cannot produce the desired

configuration since the rows of Cu–O–Cu–O chains have to change by about 90° in orientation, whereas the reset of the crystal structure does not change topologically and is only slightly sheared. Figure 10a represents the CuO_{1-x} layers of a crystal containing a ledged twin interface. A twinning dislocation is present at the position of the ledge. The magnitude of the twinning vector, as deduced from the schematic of Fig. 10b, is given by

$$t = \sqrt{a^2 + b^2 - 2ab\cos q} = b - a$$

since q is small ($\cos q = 1$). In order to restore the ortho I structure, the propagation of the twinning dislocations must be accompanied by 'reshuffling' of the oxygen atoms in the CuO layers along the core of the dislocation in the sense indicated by the curved arrows. The indicated shear stress in Fig. 10a induces the downward motion of the twinning dislocation. This results from the downward shift of the heavy atoms in the top part of the twin with respect to those in the lower part; the oxygen atoms, on the other hand, have to move in the opposite sense

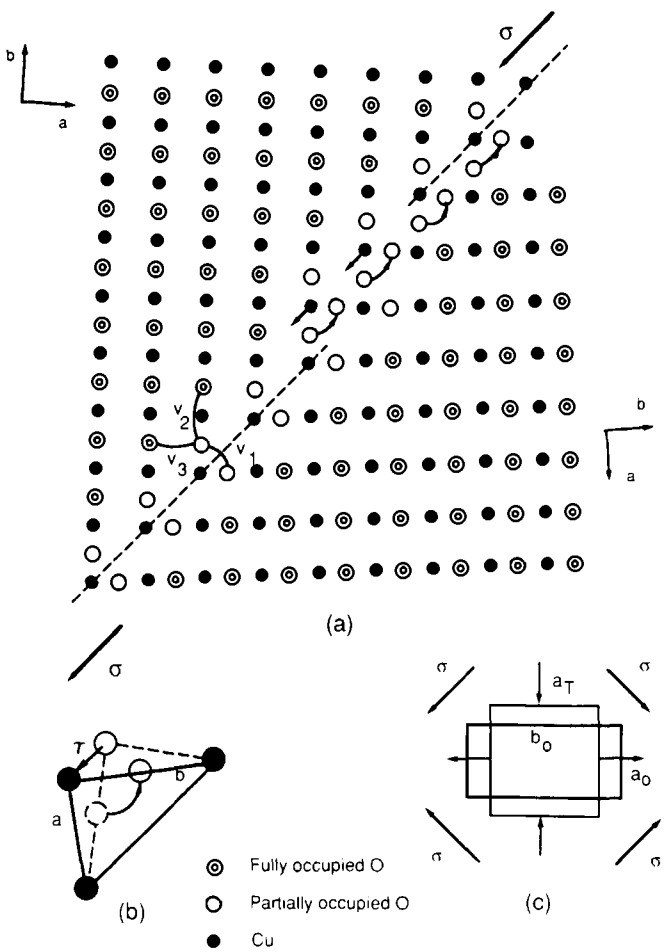


Figure 10. Schematic representation of a twinning-dislocation in $YBa_2Cu_3O_{7-x}$. (a) Atomic configuration in the CuO_{1-x} plane. (b) Atom movements during motion of the dislocation. (c) Stress pattern resulting from the orthorhombic deformation caused by ordering.

over $\frac{1}{2}[1\bar{1}0]$, as indicated by curved arrows. The resulting type of motion has been termed 'synchro-shear' motion [48]. The propagation of one twinning dislocation along the interface changes the width of the twin band by d_{110}. The rapid motion of the ledges is presumably facilitated by the presence of oxygen vacancies along the twin interface. If the boundary was highly oxygen deficient and consisted of a narrow strip of $YBa_2CU_3O_6$ structure, the reshuffling of the oxygen atoms would not even be necessary as long as such a strip is present.

It was found experimentally that twinned crystals can be 'detwinned' by the application of a uniaxial compressive stress along the [100] or the [010] directions of the perovskite lattice at a temperature of 400–450°C under a suitable partial pressure of oxygen so as to maintain the composition [49]. Similar procedures are known to be applicable to many types of crystals such as quartz, $BaTiO_3$, and ferroelastic and ferrobielastic crystals in general. Detwinning under stress can be understood on the basis of the dislocation model presented in Fig. 11. The application of a compressive stress along the *a* (or *b*) direction results in a resolved shear stress on the (110) and $(1\bar{1}0)$ coherent twin planes, which are also the glide planes of the twinning dislocations; the latter can thus move conservatively on such planes. Moreover, the compressive stress causes a bias on the sense of jumping of the oxygen atoms. The a_0 lattice parameter, which measures the Cu–Cu distance, is smaller than the b_0 lattice parameter, which equals the Cu–O–Cu distance. Jumping of an oxygen atom in the sense indicated in Fig. 11 will thus transform a b_0 distance into an a_0 distance. As a result, jumps

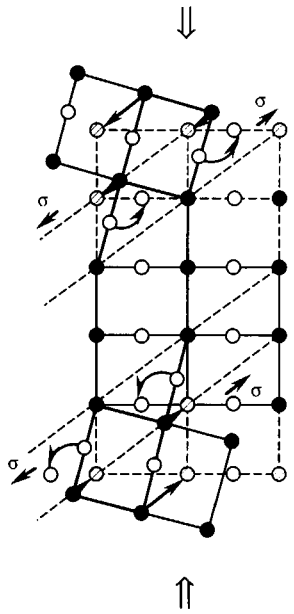


Figure 11. Schematic representation of the atomic movements during 'detwinning' under compressive stress.

which transform a Cu–O–Cu arrangement oriented along the compression axis into a Cu–Cu arrangement are energetically favored since they relieve the applied stress by shortening the specimen. These jumps are precisely those required for the synchro-shear mechanism to operate. Under the uniaxial stress the twin walls limiting a *b* domain (i.e., a domain with the *b* axis along the stress axis) will move towards one another by the combined conservative motion of twinning dislocations and oxygen jumps. Conversely, the twin walls limiting an *a* domain will move apart. The rate-determining step in the process is the short-range diffusion of oxygen along the twin boundary. Increasing the temperature to 400–450°C is therefore necessary to achieve a reasonable oxygen mobility and hence a measurable rate of the twin boundaries.

(001)-Type Defects and the Formation of $YBa_2Cu_4O_8$ and $Y_2Ba_4Cu_7O_{15}$

It is well known that deviations from the ideal composition can be incorporated in a crystal in a variety of ways, either as point defects or as extended defects. In the latter case there are still different possibilities:

(i) nonconservative stacking faults or antiphase boundaries, in chemical terminology called shear planes;
(ii) nonconservative or chemical twinning; and
(iii) precipitates limited by Frank loops.

In $YBa_2Cu_3O_{7-x}$, deviations from the stoichiometric oxygen composition are accommodated mainly as point defects, essentially oxygen vacancies in the CuO layers, which have been shown to form ordered arrangements (see later). Twin boundaries may only play a limited role in accommodating oxygen vacancies.

Excess copper, however, can be accommodated under the form of CuO, as planar defects situated in the *c* plane [50–52]. A high-resolution image of such a planar defect is shown in Fig. 12, viewed along two different directions [100] and [010]. It consists of a double layer of CuO, replacing the single layer of the normal structure (Fig. 13). In the [010] zone this double layer is viewed along the Cu–O–Cu chains; it can also be described as a set of parallel strips of CuO_4 squares in (100) planes, as represented in the [100] projection of Fig. 13. Simulated HREM images for different defocus values and different

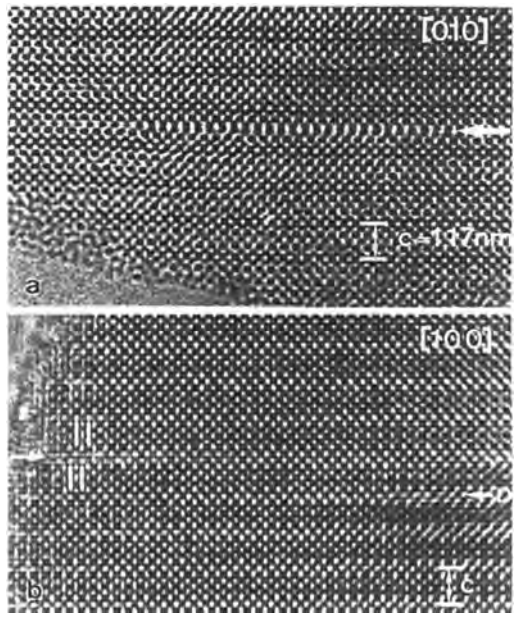


Figure 12. HREM images of planar defects in $YBa_2Cu_3O_{7-\delta}$ associated with an excess of CuO. (a) [010] view; note the Frank dislocation bordering the planar defect. (b) [100] view; note the lateral shift of $\frac{1}{2}$[010] of the two structure parts on either side of the interface.

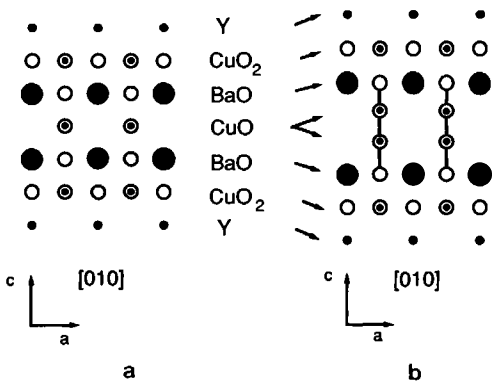


Figure 13. Atomic models of (001) defects in $YBa_2Cu_3O_{7-\delta}$. (a) The regular CuO_2–BaO–CuO–BaO–CuO_2 stacking. (b) [010] view of the planar defect. (c) [100] view of the planar defect.

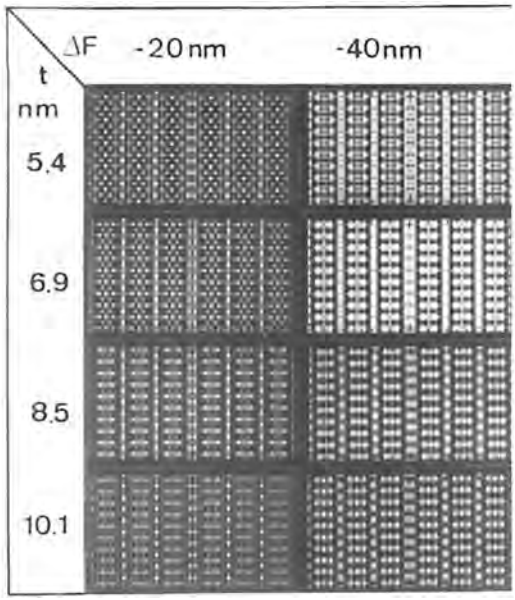


Figure 14. Image simulation of the planar defect along the [010] direction, for different thicknesses and defocus values of −20 and −40 nm.

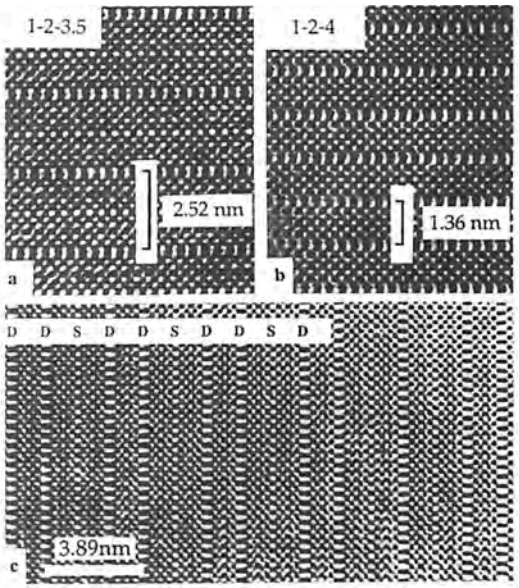


Figure 15. [010] HREM images of different YBCO polytypoids. (a) The perfect 2-4-7 structure. (b) The perfect 1-2-4 structure. (c) A more complex polytypoid, showing a mixture of double and single CuO layers.

crystal thicknesses obtained by means of the direct space algorithm are reproduced in Fig. 14, and confirm the model.

These observations have led to the development of a superconducting mixed layer YBCO family containing regular arrangements of single and double CuO layers. Members of this family are the so called 1-2-4 phase ($YBa_2Cu_4O_8$) [53, 54] and the 2-4-7 phase ($Y_2Ba_4Cu_7O_{15}$) [55]. 1-2-4 and 2-4-7 YBCO can be regarded as members of a series of 'shear structures' derived from the 1-2-3 basic structure by the periodic insertions of supplementary CuO layers. Representing the number of 1-2-3 unit cells separating these supplementary CuO layers by an integer, the 1-2-4 and 2-4-7 structures are represented by the stacking symbols $1\bar{1}$ and $2\bar{2}$ [56, 57]. The minus sign is introduced to indicate that the presence of a double CuO layer causes a lateral offset between successive 1-2-3 blocks over $\frac{1}{2}[010]$. High-resolution images along the [010] direction of the perfect 1-2-4 and 2-4-7 structures are shown in Fig. 15a and b. The double-chain layers are characterized by intense rows of elongated double white dots. Changing the initial composition does not alter the stoichiometry of the resulting material; however, to some extent different phases such as $2\bar{3}$ and $3\bar{3}$ are stabilized over limited areas. Figure 15c shows an example of a complex polytypoid over a limited region. Apart from this series of phases being a mixture of 1-2-3 and 1-2-4 where periodically double $(CuO)_2$ layers are introduced, a different series of copper-rich phases was discovered, in which successive perovskite blocks are separated by triple $(CuO)_3$ layers [58]. An example is reproduced in Fig. 16, together with a schematic inset where the CuO layer

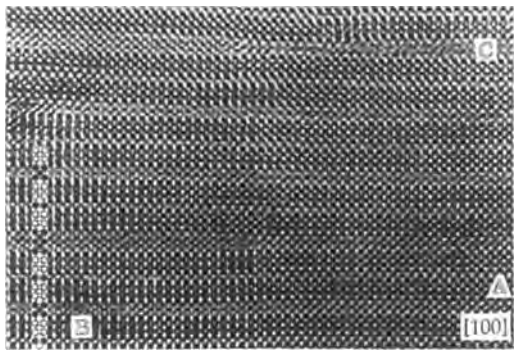


Figure 16. HREM image of YBCO, showing the local presence of triple $(CuO)_3$ and even $(CuO)_5$ layers; in the inset the CuO layers are indicated in black.

configuration is represented by black dots. In some parts of the crystal even isolated CuO multilayers consisting of as many as five CuO layers have been observed in a matrix of essentially 2-4-7 material [56, 57].

Copper Substitution in $YBa_2Cu_3O_7$ Superconductors

Although it seems well established that superconductivity is associated with the CuO_2 layers, doping of the CuO layers seems to play an important role in determining the concentration and the type of carriers. Also, the critical temperature T_c is related to the composition of the CuO layer. It could for instance be shown that the 60 K plateau in the curve relating T_c to the oxygen content in this layer is associated with the presence of the $2a_0$ structure, that is, with the material with ideal composition $YBa_2Cu_3O_{6.5}$ [59–61]. It was attempted to replace copper in the CuO layers, either completely or partly, by other metallic ions such as iron, cobalt, zinc, gallium. Also complex ions such as CO_3^{2-} or SO_4^{2-} were incorporated in the CuO layer, carbon and sulfur occupying copper sites in the CuO chains. Iron substitution is found to induce a 'tweed' texture, the geometry of which could be assessed by assuming that iron occurs in the CuO layers in two types of oxygen coordination: octahedral and tetrahedral. The tetrahedrally coordinated iron ions induce microtwinning by the nucleation of CuO chains along mutually perpendicular directions, which causes the tweed microtexture [62]. Cobalt substitution in small concentrations has a rather similar aspect [63].

Oxygen-Deficient $YBa_2Cu_3O_{7-\delta}$

From a microstructural point of view, $YBa_2Cu_3O_{7-\delta}$ is an interesting compound. It allows a large number of oxygen vacancies ($0 \leqslant \delta \lesssim 1$) in the oxygen sublattice. Depending on their concentration, these vacancies tend to occur in various ordering schemes [33, 46, 64–83]. The $2a_0$ superstructure (ortho II) in nonstoichiometric $YBa_2Cu_3O_{7-\delta}$, due to the alternation along the a_0 direction of filled and empty CuO chains in the CuO layers, was also discovered by electron diffraction and electron microscopy [64, 79, 80]. The oxygen ordering occurs within nanometer-sized domains and is generally highly faulted; therefore, it is difficult to detect by X rays or neutrons. On the other hand, these small dimensions make them eligible for flux pinning centra.

Another important feature of the YBCO compound is its susceptibility to allow elemental substitutions on most of its sublattices, without substantially altering the structure, but inducing quite different physical and chemical properties. We will focus here on those aspects where electron microscopy and electron diffraction have played important roles in

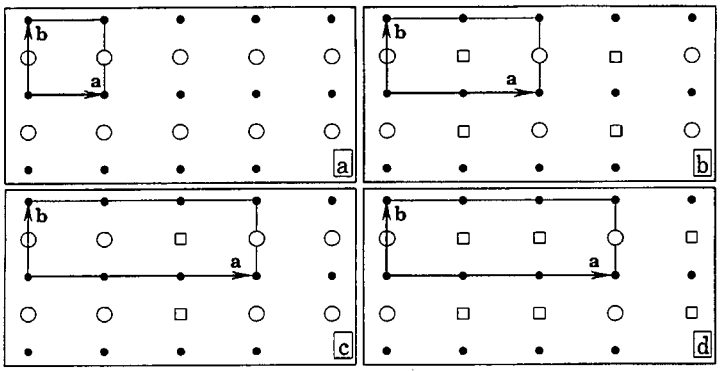


Figure 17. (a) Schematic representation of the CuO plane ortho I ordering. (b) Model of the ortho II ordering. Along a CuO and Cu vacancy, chains alternate. Representations of ordering leading to tripled a-parameters: (c) the ortho III phase with an oxygen content of 6.667; (d) a symmetric phase where roles of vacant and filled chains are reversed, which occurs at oxygen contents of 6.333.

elucidating the structure and bridging the gap between physical properties and structural data.

At room temperature and atmospheric pressure, within the range $0 \leqslant \delta \lesssim \delta_t$, the structure of $YBa_2Cu_3O_{7-\delta}$ is orthorhombic and within the range $\delta_t \lesssim \delta \leqslant 1$, it is tetragonal. Values for δ_t vary around 0.65 [68, 84–88]. In experiments where δ is fixed, the transition temperature T_t increases linearly with δ [89]. Reported values for the orthorhombic to tetragonal transition temperature range between 110°C at $\delta = 0.67$ and 680°C at $\delta = 0.34$.

In the tetragonal phase, the oxygen atoms in the CuO layer are distributed quasi-randomly, with an equal occupation of the O(1)- and O(5) sublattices. The sublattices consist of the oxygen sites between every two copper atoms in the plane, the O(1) sites lying along one direction (defined as ***a***), the O(5) sites along the other (Fig. 17). The formation of short CuO segments along both basic directions of the CuO chain plane is likely to cause local orthorhombic strains that lead to the pretransition tweed texture.

The presence of short-range ordering of oxygen and vacancies is revealed in electron diffraction patterns by the occurrence of diffuse intensities. The diffuse streaks observed in the [001] zone diffraction pattern of Fig. 18 are due to the presence of irregularily spaced, short segments of CuO chains. Short-range order is commonly observed in highly oxygen-deficient samples and can be induced in oxygen-rich materials by a thermal disordering by an in situ heating treatment.

The orthorhombic phase that corresponds to $YBa_2Cu_3O_7$ ($\delta = 0$), is termed 'ortho I'. The CuO plane in the ortho I phase has all O(1) sites filled and all O(5) sites vacant. CuO chains run along ***b*** (Fig. 17a). As a result, for the ortho I phase $a_0 \lesssim b_0$. A second phase, termed ortho II, has every other ***b***-oriented chain evacuated

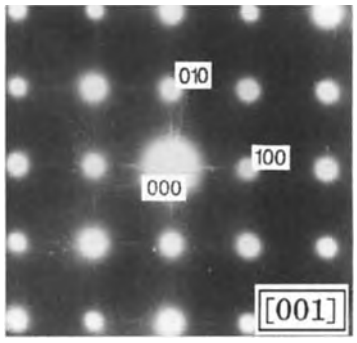


Figure 18. Electron diffraction pattern along the [001] zone, in a tetragonal sample (from the observation of unsplit spots). In situ heating and cooling causes disorder reflected in the presence of diffuse scattering along the basic directions.

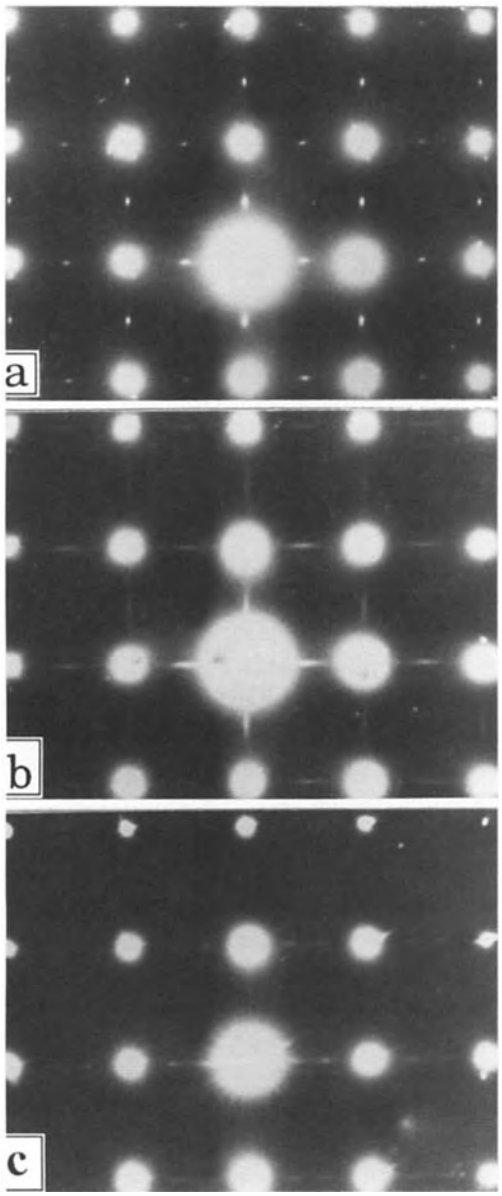


Figure 19. (a) [001] zone diffraction pattern with sharp superstructure reflections at positions $h + \frac{1}{2}k\,l$ appearing in well-ordered ortho II material. Since the material is twinned, spots seem to appear along both basic directions. (b) [001] zone diffraction pattern with streaked superstructure intensity, corresponding to domains that are short-range ordered along ***a***, occurring at oxygen contents at the slopes above and below the ortho II plateau. (c) [001] zone diffraction pattern with spots at positions $h + \frac{1}{3}k\,l$ and $h + \frac{2}{3}k\,l$ due to an ortho III structure occurring in samples with oxygen contents around 6.8.

of oxygen and has an ideal composition $YBa_2Cu_3O_{6.5}$ ($\delta = 0.5$) (see Fig. 2b). The oxygen content range in which the ortho II phase is observed stretches over $0.3 \lesssim \delta \leqslant 0.5$. The new unit mesh that can be defined on the ordered $CuO_{1-\delta}$ plane has dimensions $a_{II} = 7.668\,\text{Å} \approx 2a_p$, $b_{II} = 3.869\,\text{Å} \approx b_p$ (based on Cara et al. [64]). The ortho II phase is easily recognized in [001] electron diffraction patterns (Fig. 19); it produces superstructure spots at positions $h + \frac{1}{2}kl$, characteristic for a doubling of the a parameter. Often, however, in such patterns the b parameter also seems doubled by the appearance of spots at positions $hk + \frac{1}{2}l$ as well. These spots are due to their basic (110) twinning of the orthorhombic matrix, the resulting diffraction pattern (Fig. 19) being the overlap of the two identical patterns, with a rotation difference of about 90°. It was shown by Van Dyck et al. [90] that a disordered array of line defects gives rise to Lorentzian-shaped diffraction spots, with a position associated with the mean spacing of the scattering elements. When chains in the CuO plane are irregularly spaced with a mean spacing of $2a_p$, Lorentzian-shaped diffraction spots at positions $h + \frac{1}{2}kl$ are expected. When the average periodicity increases, the superstructure reflections will become more elongated (Fig. 19b). A second factor influencing the superstructure reflection shape is the ortho II domain shape, small domain dimensions leading to elongated reflections. Intensity and sharpness of the superstructure reflections can thus be considered as indices of the domain size and the ordering quality of the phase. Whereas

an ideal ortho II domain has an oxygen deficiency of $\delta = 0.5$, the strongest and sharpest ortho II electron diffraction intensity occurs at an oxygen deficiency of only $\delta = 0.4$ [91]. The domain size and the ordering quality are thus apparently maximized at $\delta = 0.4$.

Dark field images allow the visualization of the oxygen-ordered domains. The images show that the ortho II domains are elongated or lenticular in shape, with the short axis along the ***a*** direction. The largest ortho II domains (4×20 nm) and a maximal ortho II volume occur at oxygen deficiencies of $\delta = 0.4$. For the $\delta = 0.4$ sample of Zhu et al. [67], from measurements of the ortho II peak surface in a [100] electron densitogram, the ortho II volume fraction was an estimated 50%, in accordance with our results. Samples showing elongated ortho II diffraction spots, correspondingly show ortho II domains in dark field images (Fig. 20b) with reduced dimensions along ***a*** (size: 1×20 nm).

Under high-resolution conditions, imaged along the [100] direction, the ordered domains can be easily visualized. Such images also allow deduction of the stacking of the vacancy rows in successive $CuO_{1-\delta}$ planes along the c axis (Fig. 21). The normal stacking of these rows is vertical, but small domains of a staggered stacking of successive two-dimensional-ordered CuO planes are no exception. The ordered domains also reveal antiphase boundaries with a displacement vector $\boldsymbol{R} = \frac{1}{2}[100]_{OIIa}$. This is best visible under grazing angles along the [001] direction, and indicated by a white line, broken at the defects.

Instead of a doubling of the unit cell due to oxygen vacancy ordering along the

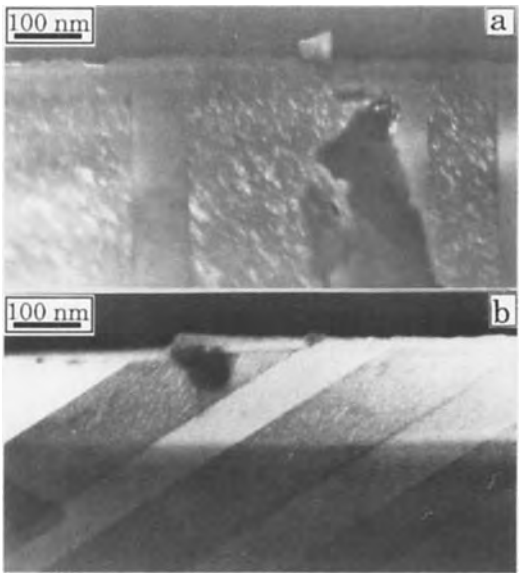


Figure 20. (a) Dark field image showing ortho II ordered domains as bright areas. The corresponding diffraction pattern is shown in Fig. 19a. (b) Dark field image corresponding to Fig. 19b. Domain size is short along ***a*** and large along the CuO chain direction ***b***. The band structure in these images is due to twinning.

b axis, one can imagine different ordering schemes, introducing tripling, quadrupling, and so on of the basic ortho I structure. Electron microscopic observations of the ortho III phase have been reported (Fig. 19c), but much less and more ill-defined than the ortho II phase [72, 92, 93]. Due to the short structural coherence length of this phase, their presence could not be detected neither by X-ray nor by neutron diffraction. The ortho III phase ideally appears at an oxygen deficiency of $\delta = \frac{1}{3}$ (or $\frac{2}{3}$), and compared to the ortho I phase, one out of three Cu_1O_1 chains is depleted of oxygen (see Fig. 17c).

Superstructures with longer periodicities of $4a_p$ and $5a_p$ have also been observed by electron microscopy [94, 95]. These

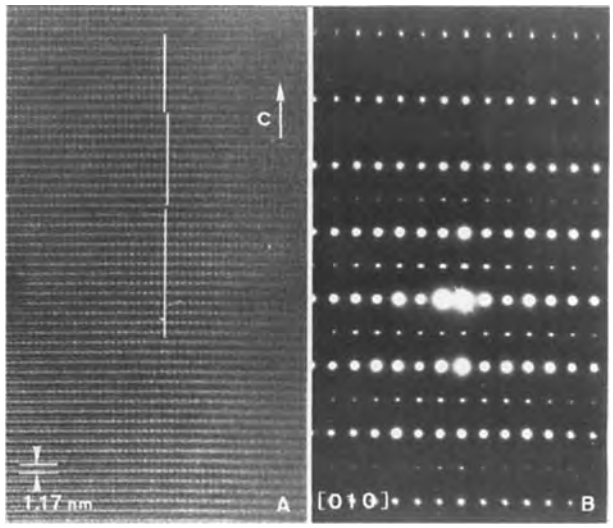


Figure 21. High-resolution images along the [010] axis, of the ortho II phase. The cell doubling can be seen by the white dots that appear every $2a_p$. In this view along the chains, chains are stacked vertically along *c*. Note the presence of antiphase boundaries with displacement vector $\boldsymbol{R} = \frac{1}{2}[100]$.

phases appear on very local scales (a few unit cells wide), and therefore have only been observed by means of HREM. Identifying these phases as (Cu–O) chain ordered phases, they should probably be considered as metastable phases at oxygen contents intermediate to that of the stable ortho I, ortho II, and ortho III phases. Theoretical studies seem to support this conclusion [96, 97].

8.2.3.2 Superconducting Materials and Modulated Structures

Modulated structures, that is, superstructures, commensurate or incommensurate, built on a basic structure, occur frequently in organic chemistry, and this is not different in the family of high-T_c superconducting compounds. We have already discussed the vacancy ordering in YBCO, which introduced doubling, tripling, and so on of the basic unit YBCO unit cell along the *b* axis. We will briefly review here the deformation-modulated structures in the bismuth family (BSCCO) of superconducting compounds and the modulated structures in the thallium- and mercury-based materials by the introduction of carbonate groups. We will also briefly comment on the relationship (or absence of it) between superconductivity and the presence of modulated structures. In less detail we will describe the unusual modulation in the copper-free superconductor $Ba_{1-x}K_xBiO_3$ and the interface-modulated structure in oxygen-deficient La_2CuO_{4-x}.

$Bi_2Sr_2Ca_nCu_{1+n}O_{6+2n}$

It is well established now that the superconducting transition temperature for the $Bi_2Sr_2Ca_nCu_{1+n}O_{6+2n}$ and $Tl_2Ba_2Ca_nCu_{1+n}O_{6+2n}$ compounds depends upon the composition of the material, that is, upon the value of *n*. The critical temperature for the $Bi_2Sr_2Ca_nCu_{1+n}O_{6+2n}$ compounds increases with increasing *n* value from 20 K for $n = 0$ to 80 K for $n = 1$ and to 110 K for $n = 2$ [98–100]. For the thallium

compounds these changes are 80 K for $n = 0$, 110 K for $n = 1$, and 125 K for $n = 2$ [101].

In the bismuth compounds the basic structure is further modulated. This modulation is incommensurate for all compounds but varies with composition, oxygen content, and doping with foreign elements. However, no visible temperature dependence of the modulation has been observed between 20 and 750 K. Since the modulation is a structural aspect, occurring at the atomic level, the most suitable technique to image and investigate it is TEM. Other imaging techniques lack the resolution or they only give information about the surface effects (e.g., AFM or STM). The modulation was first studied by TEM and only later by diffraction techniques such as X-ray or neutron diffraction.

The pure $Bi_2Sr_2Ca_nCu_{1+n}O_{6+2n}$ compounds for different values have been studied by a large number of groups; their HREM results can be found in the literature [102–110]. The effect on the modulation of substituting several elements, such as lead for bismuth, lanthanum for strontium, or yttrium for calcium, is also well documented [111–120].

The most illustrative section in reciprocal space for the description of the modulated structure is $(110)^*$, which is reproduced for the pure BSCCO compound in Fig. 22 for $n = 0$ and 1. Every reflection of the type $00l$ or $11l$ acquires a linear sequence of several satellites. These satellites are situated in the $(110)^*$ plane, but they are not oriented along a simple crystallographic orientation; moreover, the orientation as well as the length of the modulation vector is slightly variable. The wave vector can be modified by partially substituting Sr^{2+} by La^{3+}; details

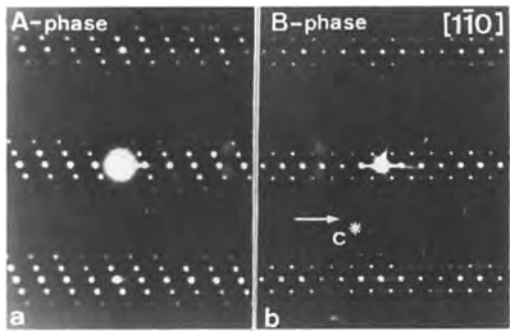


Figure 22. [1$\bar{1}$0] electron diffraction patterns in $Bi_2Sr_2Ca_nCu_{1+n}O_{6+2n}$ showing the incommensurate modulation for (a) the $n = 0$ compound (A phase) and (b) the $n = 1$ compound; the modulation is clearly different.

about these experiments as well as structural models and speculations about the modulation can be found in Zandbergen et al. [120]. In the undoped $n = 0$ material the modulation is roughly along $[115]^*$, but the wavelength is still incommensurate and is of the order of 2.4–2.5 nm. The high-resolution images along the [110] zone clearly reveal the modulation as well as the lattice distortions associated with it (Fig. 23). The (001) planes in the image

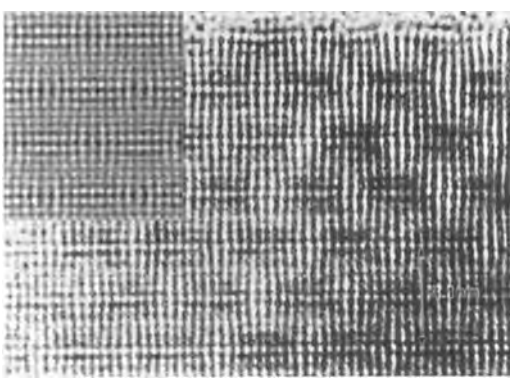


Figure 23. HREM of the $n = 1$ phase of BSCCO, viewed along the [110] direction. Both the (001) and (110) planes exhibit a wavy character. The computed image, for a model exhibiting an excess oxygen shows good correspondence with the experimental micrograph.

are clearly buckled; the lines of more intense white dots can be interpreted as BiO layers, and the variation in spacing along [110] is appreciable. The $n = 1$ phase and $n = 2$ phase produce a similar diffraction pattern (see Fig. 22b); the satellites are also located in the $(110)^*$ plane, and the wavevector $\boldsymbol{q}$ can be chosen along $[110]^*$ and has a wavelength of 2.5–2.6 nm. In the direct image the distortion of the (001) layers is clearly strongest for the $(BiO)_2$ double planes. The observed spacing changes in the BiO–BiO layer distance is of the order of 0.07 nm for the $n = 1$ compound; for $n = 2$ this distortion seems to be less pronounced. The modulation reduces the symmetry of the structure to orthorhombic with point group *mmm*. Such a description of the modulated structure is more elegant in a four-dimensional space; taking the *a* and *b* axis along $[1\bar{1}0]$ and [110], respectively with a length of 0.54 nm and the modulation along the *b* axis one observes systematic extinctions for *hklm* when $h + l + m = 2n + 1$, which suggests the space group to be $L^{Bmmm}_{1\,\underline{1}\,1}$.

Over the years there has been a lot of speculation and a lot of experiments performed to find out the exact reason for the modulation, as well as its relationship with the superconducting properties. Some of the possible reasons for the modulation are: the partial absence of strontium atoms, the partial substitution of bismuth by copper or strontium, the incorporation of extra oxygen, a change in the position of the lone pairs of bismuth, and a lattice mismatch between the BiO double layer and the perovskite layer. These possibilities are discussed in more detail and compared with the experimental evidence by Zandbergen et al. [120].

At this moment the most widely accepted model combines several aspects of the originally proposed model. The main reason for the modulation is the structural misfit between the perovskite slab and the BiO_2 double layer; this misfit introduces strains in the double layer, which because of the asymmetrical arrangement of the electronic configuration (lone pair) of bismuth allows it to be deformed. This deformation then creates the possibility of introducing extra oxygen. This interpretation has also been confirmed by X-ray and neutron experiments [121–126]. Calestani and co-workers [127, 128] have done clear experiments to replace calcium by yttrium and/or bismuth by lead, and following the modulation behavior as well as the superconducting properties. Their conclusion is that superconductivity can occur even in unmodulated material, whereas modulated materials can easily be obtained which are nonsuperconducting. Pham et al. [129] also conclude that superconductivity and structural modulation are not related; also, they find that a modulated structure can exist in pure BSCCO material without any oxygen excess.

$Tl_2Ba_2Ca_nCu_{1+n}O_{6+2n}$

The thallium compounds with compositions $Tl_{2-x}Ba_2Ca_nCu_{1+n}O_{6+2n}$ exhibit the phenomenon of 'mixed layer polytypism', which consists of the occurrence of different stacking sequences of similar lamellae with different thicknesses so as to lead to structures with different ideal chemical compositions. If instead of a double TlO layer only a single (or eventually a triple) TlO layer is present, no offset is observed between successive

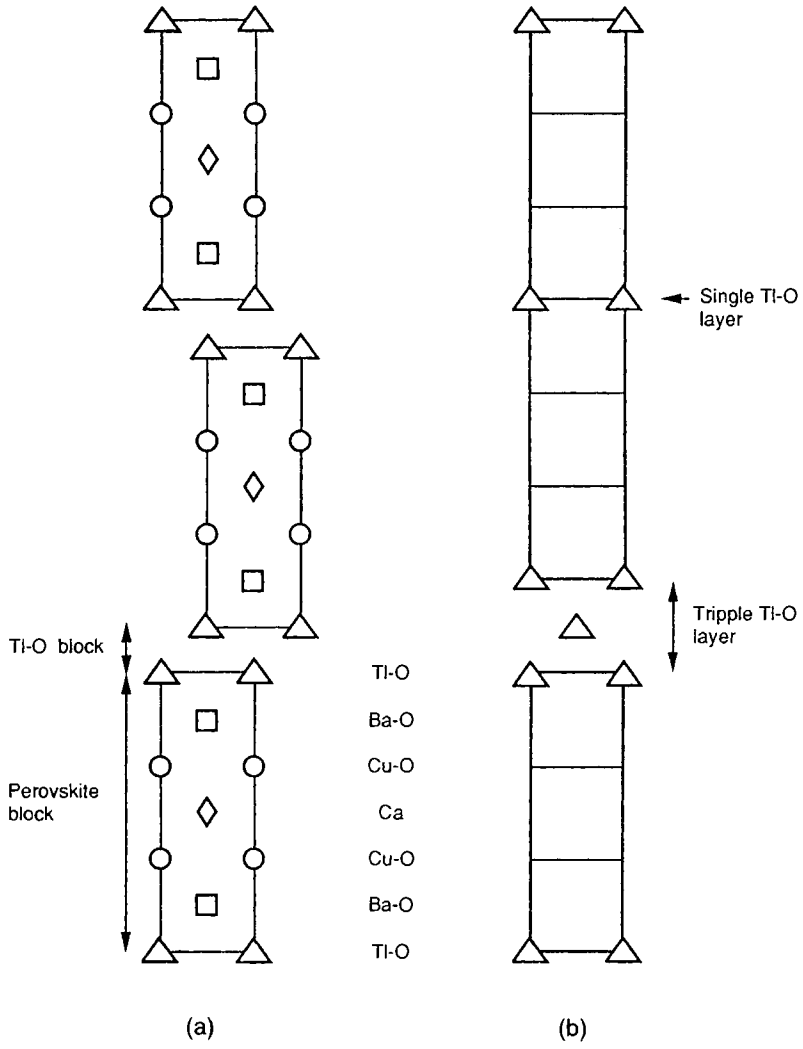


Figure 24. Schematic representation of thallium-based superconducting compounds. (a) The regular 2212 configuration, with double TlO layers. (b) A defect configuration with single or triple TlO layers.

perovskite blocks; this is shown schematically in Fig. 24. Such structures having a monolayer of TlO have been discovered by Parkin et al. [130]. In a specimen with thallium deficiency one can observe stacking sequences with very large periods such as the one reproduced in Fig. 25a. The diffraction pattern of such an area apparently exhibits streaks along the c^* direction, but in reality these streaks can be resolved as sharp spots, suggesting that a regular stacking sequence with a very long period is present (Fig. 25b). To determine the stacking sequence in such a long-period compound, HREM is a unique tool. An example in which all atom columns are clearly resolved is reproduced in Fig. 26. The period is 21 nm and the stacking sequence is indicated. It should further be noted that between a number of successive perovskite blocks an offset is observed whereas this is not the case

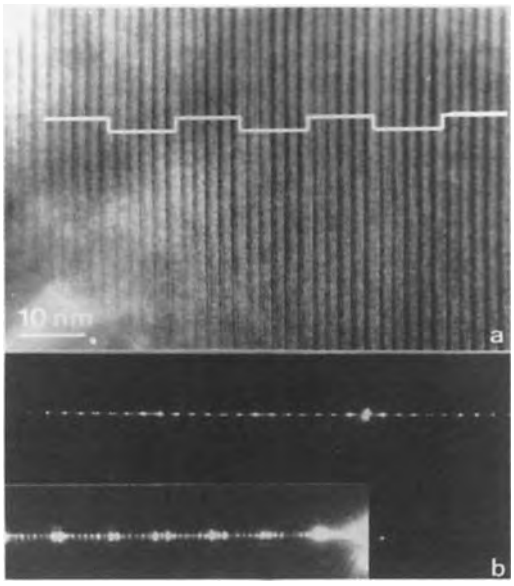


Figure 25. (a) Low magnification of a thallium long-period mixed-layer polytype; the periodicity has been indicated. (b) Electron diffraction pattern of a long-period polytype showing almost continuous streaking. The enlarged inset, however, reveals a dense arrangement of sharp diffraction spots.

between other successive perovskite blocks. This indicates that double as well as single TlO layers occur. This has also been indicated in Fig. 26 where a single and a double bar are used to indicate, respectively, a single and a double TlO layer. A striking feature apparent from this figure is the fact that the sequence of blocks, although complicated, repeats over large areas and that, moreover, also the sequence of single and double layers of TlO repeats with the same regularity. The unit cell of the resulting mixed-layer polytype is body centered. More details, and a convenient way to describe such complex polytypes, can be found in the literature [131, 132].

Long-Period Shear Structure in La_2CuO_{4-x} [133, 134]

The undoped La_2CuO_4 crystallizes in an orthorhombically deformed K_2NiF_4-type structure with lattice parameters $a = 0.5355\,\text{nm}$, $b = 0.5441\,\text{nm}$, and $c = 1.353\,\text{nm}$. The orthorhombic distortion introduced below 530 K is due to a correlated tilting of the CuO_6 octahedra, which are assumed to be rigid and corner linked.

In oxygen-deficient material, however, a shear structure is formed to accommodate the local nonstoichiometry. Evidence is shown in Fig. 27, which is a high-resolution image along the perovskite *a*

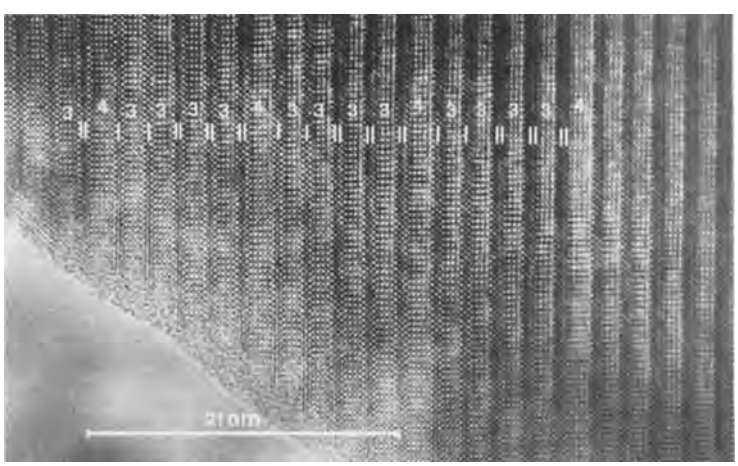


Figure 26. High-resolution image of the long-period polytype of Fig. 25a; the repeat unit of 21 nm is indicated together with the building sequence. The perovskite blocks are indicated by the numbers 3 ($n = 3$) or 4 ($n = 4$), while double and single TlO layers are indicated by double and single lines, respectively.

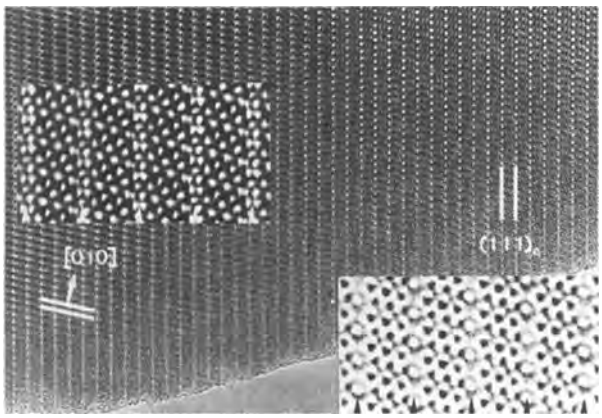


Figure 27. Large area of the La_2CuO_{4-x} shear structure viewed along the $[100]_p = [101]_o$ zone; insets are magnified images at two different thicknesses. The ion columns are imaged as bright dots.

axis, that is, along the orthorhombic [101] axis. Vertical crystallographic shear planes, parallel to the (111) planes of the orthorhombic unit cell, show up prominently in this image. The corresponding diffraction pattern is reproduced in Fig. 28a; the rows of closely spaced reflections are perpendicular to the family of interfaces revealed in Fig. 27, and their spacing is consistent with the observed spacing between successive interfaces. Exploring carefully reciprocal space indicated that all reflections can be indexed in a new monoclinic unit cell related to the basic perovskite lattice parameters by the relation

$$\begin{matrix} \boldsymbol{a}_m \\ \boldsymbol{b}_m \\ \boldsymbol{c}_m \end{matrix} = \begin{matrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 2 & 0 \\ 7 & 0 & -2\epsilon \end{matrix} \begin{matrix} \boldsymbol{a}_{1,p} \\ \boldsymbol{a}_{2,p} \\ \boldsymbol{c}_p \end{matrix}$$

where ϵ is a small quantity accounting for the offset in the b_0 direction along the shear planes. The approximate lattice parameters of the new interface modulated phase are

$$a_m = 0.66\,\text{nm}$$
$$b_m = 0.76\,\text{nm} \qquad b = 103°$$
$$c_m = 2.66\,\text{nm}$$

From the reflection conditions (l even for $h0l$- or $00l$-type reflections) and taking into account the possibility of double diffraction in electron diffraction, one obtains the space group $P1c1$ or $P12/c1$.

More structural information, however, can be obtained from the high-resolution images, such as that of Fig. 29. This image was obtained from a very thin part of the specimen where the heavy ion columns are imaged as dark dots; the shear planes which are less obvious here are vertical. The correspondence with the crystal structure is indicated in the right bottom

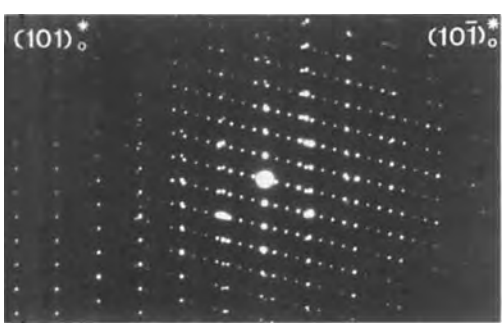


Figure 28. Diffraction pattern of the La_2CuO_{4-x} shear structure across a twin boundary showing the superposition of the $(101)^*$ section, not showing the structure (left part) and the $(\bar{1}01)^*_o$ section exhibiting rows of superstructure spots.

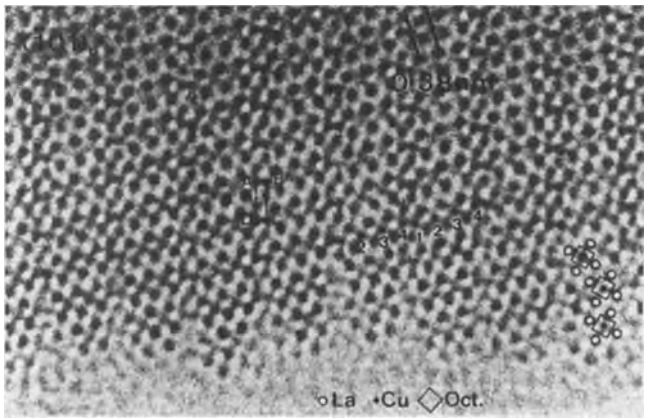


Figure 29. High-resolution image of the long-period shear structure along the a_p zone. The heavy ion columns are imaged as dark dots, the oxygen columns remain invisible. The relation with the basic La_2CuO_4 structure is indicated.

corner. Along the shear planes the projected separation of copper ion columns, such as those numbered 1 and 4, is about $\frac{1}{2}a_p$, which is consistent with the projected separation of the centers of edge-sharing CuO_6 octahedra. This projected structure is represented in Fig. 30.

The proposed model implies that the oxygen loss and the accompanying crystallographic shearing take place above the tetragonal–orthorhombic transition temperature of 530 K, thereby forming a shear structure with the space group $P12/c1$. On cooling, tilting of the

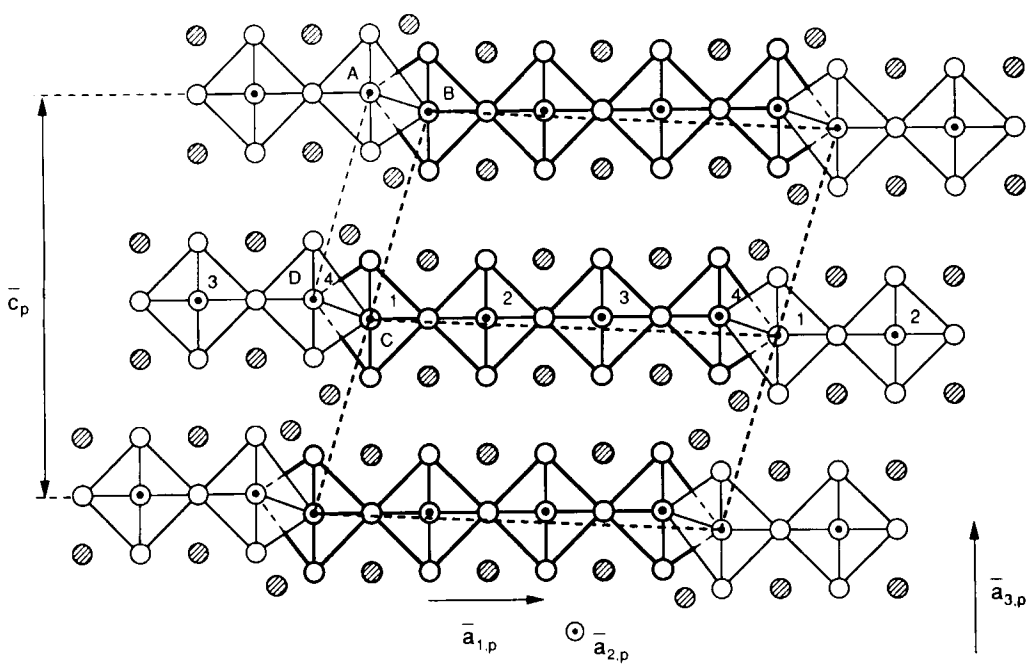


Figure 30. Schematic view of the La_2CuO_{4-x} shear structure along the zone parallel with a_p; the shear planes are formed by edge-shearing CuO_6 octahedra; the tilt of the octahedra has not been represented.

octahedra takes place [133], which is imposed by the edge and corner sharing of the octahedra. The twofold axis is lost and the space group now becomes $P1c1$, which is a subgroup of $P12/c1$, exhibiting the same diffraction conditions.

This is clearly an example of a long-period interface modulated structure where one can determine a new structure exclusively from:

(a) the structure of the unmodulated phase;
(b) electron diffraction of the modulated phase, yielding the space group; and
(c) direct imaging by HREM to determine the displaced atom positions.

Incommensurate Deformation Modulation in $Ba_{1-x}K_xBiO_3$

The overwhelming majority of the new high-T_c superconductors do contain copper; however, $Ba_{1-x}K_xBiO_3$ with $x \approx 0.4$ shows bulk superconductivity below 30 K. Its basic structure is that of the perovskite ABO_3 with A = Ba or K, and B = Bi. It consists of a framework of corner-linked oxygen octahedra. Depending on composition and temperature, rigid tilt of the BO_6 octahedra, combined with a deformation of the octahedra ('breading mode distortion'), give rise to a number of structural phase transitions. More details on these phase transitions and on their relationship to the superconducting properties are given by Verwerft et al. [135].

A structural modulation, incommensurate with the basic lattice, was first reported by Chaillout and Remeika [136] in oxygen-deficient $BaBiO_{3-x}$; Pei et al. [137, 138] observed a similar modulation in $Ba_{1-x}K_xBiO_3$ for all the nonsuperconducting compositions. They showed that the modulation wavelength varies systematically with the potassium content. The interpretation of the modulation has been the subject of several speculations. According to some [38] it is to be attributed to a vacancy ordering, others [137, 138] associate the modulation with the presence of a charge density wave, while Hewat et al. [139] proposed a model involving partial ordering of Bi^{3+} versus Bi^{5+}, accompanied with bismuth or oxygen displacements correlated with the Bi^{3+} lone pair ordering.

Electron irradiation experiments were carried out for accelerating voltages between 40 and 400 kV. Below 100 kV no atom replacements, not even for the light oxygen, are expected. Typical behavior for the investigated $Ba_{1-x}K_xBiO_3$ samples under the electron beam is shown in Fig. 31. After a minimal electron dose, no satellite reflections are observed (Fig. 31a); superstructure reflections appear after

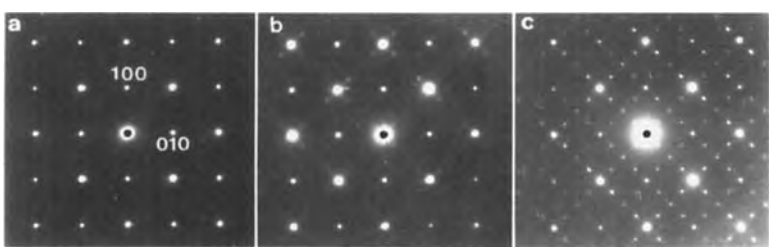


Figure 31. Electron irradiation effect observed in a [001] zone axis orientation of $Ba_{1-x}K_xBiO_3$ (a) after a few seconds of observation. No satellite reflections are observed. (b, c) On irradiating the sample with a defocused electron beam, satellite reflections become visible, and their intensity increases with irradiation.

longer irradiation times (Fig. 31b and c). Care is taken to avoid heating of the sample, so that the modulated structure is to be associated with an electron irradiation effect. Around each basic spot, four satellite reflections are observed in the [001] zone axis diffraction pattern. The wavevectors corresponding to this modulation can be written as $\boldsymbol{q}_1 = (\frac{1}{4} - \epsilon)[110]^*$ and $\boldsymbol{q}_2 = (\frac{1}{4} - \epsilon)[1\bar{1}0]^*$. Actually, in three-dimensional reciprocal space one finds that every basic reflection is surrounded by 12 satellite reflections along all $\langle 110 \rangle$ and $\langle 1\bar{1}0 \rangle$ directions. High-resolution micrographs show that the modulated structure is broken up into microdomains with an average size of 10 nm (Fig. 32). In each of these domains the modulation is described by a single wavevector. The electron diffraction patterns such as Fig. 31c are therefore to be interpreted as a superposition of different microdomains.

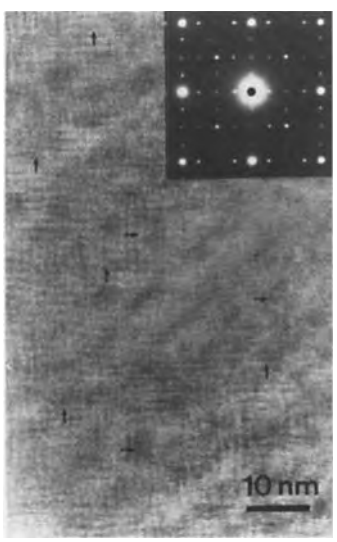


Figure 32. High-resolution micrograph evidencing the existence of microdomains. In each of these domains, the modulation is single wavevector, the modulation wavevector directions are schematized by small arrows.

In each of these microdomains the modulation is described by a single wavevector, which can be written as $\boldsymbol{q}_1 = (\frac{1}{4} - \epsilon)[110]^*$.

The wavelength of the modulation, that is, ϵ, varies with the potassium content. The incommensurability varies from $\epsilon = 0.049$ for $x = 0$, to $\epsilon = 0$ for $x = 0.4$.

In situ heating experiments in $Ba_{1-x}K_xBiO_3$ with the electron beam turned off caused all satellite intensities above 200°C to disappear. When the sample was then cooled down with the beam off, the satellites remained extinct; however, when irradiating the sample again the satellites reappeared at their initial positions. Note that the satellites do not disappear when the electron beam is turned off without heating.

Symmetry analysis of this incommensurate modulation can be worked out in the framework of the superspace formalism; details on the calculations are given by Verwerft et al. [135]. It is concluded that only bismuth atoms are involved in the modulation and moreover they can only be displacively modulated. The final modulation is represented by a $\sqrt{2}a \times \sqrt{2}a \times a$ unit cell; the modulation is a transverse displacive wave with wavevector parallel to the x axis and displacements parallel to the y axis. In the high-resolution image of Fig. 33a, viewed under grazing incidence, one detects a displacement of the brightest dots; the displacement amplitude can be estimated as 0.005 nm. This modulation of the bismuth atoms is represented schematically in Fig. 33b.

From this analysis, all models involving chemical ordering can be excluded; the ordering of Bi^{5+}/Bi^{3+} can also be ruled out since the displacements of the bismuth atoms are too small to result in a complete charge disproportion. In view of the fact

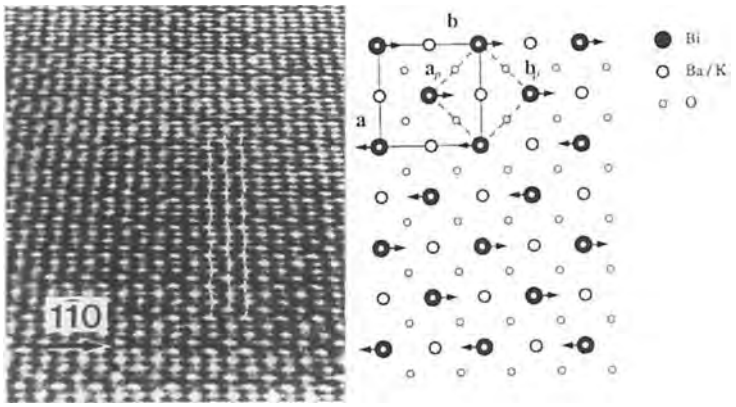


Figure 33. (a) High resolution micrograph viewed at the grazing angle, showing the displacive modulation of the brighter dots. (b) Schematic drawing of the modulation in $Ba_{1-x}K_xBiO_3$ represented in a projection along the [001] direction. Bismuth atoms are represented as black dots, barium or potassium atoms as large open circles, and oxygen atoms as small open circles.

that the modulation only shows up after electron irradiation, it might be argued that it is not a fundamental material property. This, however, is very difficult to accept because of the very low threshold value (<40 kV). Elementary collision theory shows that the maximum energy transmitted to the lightest element, oxygen, is about 5 eV; this is far too low for the creation of oxygen vacancies. We therefore propose a model of time-fluctuating charge disproportion of the bismuth atoms due to displacements from their average positions. Upon electron irradiation it is possible to induce clustering of already existing point defects (e.g., oxygen vacancies), even at very low energy transfers. These clusters can then act as pinning centers for the modulation. Upon heating above 200°C these centers may lose their pinning capability, resulting in a disappearance of the satellite reflections.

8.2.4 Holographic Interference Microscopy and Lorentz Microscopy [140]

Tonamura and co-workers at Hitachi Co. [141] have 'modified' the transmission electron microscope such that it allows direct observation of vortices with the help of electron holography (see Part I of this Handbook). In their method, magnetic lines of force emerging from the superconducting surface can be observed directly in a holographic interference micrograph or by Lorentz microscopy. The basic idea is that vortices are phase objects and consequently cannot be detected on a micrograph where the electron intensity is detected. The phase shift caused by the vortex becomes detectable, for example, by directly measuring the phase distribution in a holographic interference micrograph or by strongly defocusing the objective lens (see Part I).

By interference microscopy the magnetic lines emerging from the surface can be directly observed as contour fringes in a phase-amplified pattern, through holography. In the example of Fig. 34 a magnetic field was applied perpendicular to the lead film, which was cooled to 4.5 K; one fringe corresponds to one vortex. In addition to isolated vortices, a pair of vortices is observed, which are oriented in opposite directions and connected by magnetic force lines.

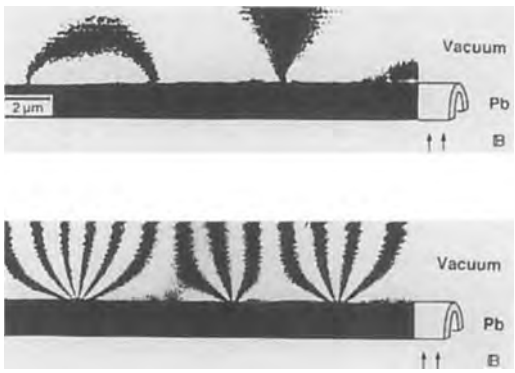


Figure 34. Interference micrographs of magnetic fields emerging from a lead-film (phase amplification: ×2). (a) Thickness 0.2 μm. (b) Thickness 1 μm. (Courtesy of Tonomura [141].)

A two-dimensional distribution of vortices can be dynamically observed by Lorentz microscopy, using a holography electron microscope. A superconducting thin film is tilted 45° to the incident beam, so that the electrons are affected by the vortex magnetic field. An external B field is applied horizontally. A Lorentz micrograph at $B = 100\,\mathrm{G}$ is shown in Fig. 35. Each spot showing a black–white contrast is an image of a single vortex. Using the same methods the vortex lattice in BSCCO ceramic superconductors

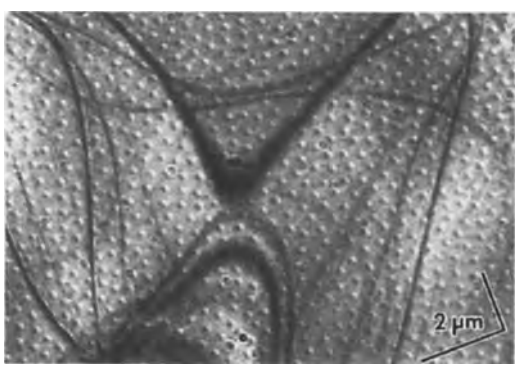


Figure 35. Lorentz micrograph of a two-dimensional array of vortices in a superconducting niobium film at 4.5 K. (Courtesy of Tonomura [141].)

can also be observed and followed as a function of temperature between 4.5 K and T_c. It is found that the image contrast gradually fades with increasing temperature.

8.2.5 Direct Observation of Vortices by Magnetic Force Microscopy

As we have seen before, STM can be applied to conventional type II superconductors to image the hexagonal Abrikosov vortex lattice as well as the electronic fine structure of a single vortex lattice [142]. The imaging of single vortices in high-T_c superconducting materials has taken a lot of effort because of a combination of short coherent length, degraded surfaces, and high mobility of the vortices. Recently, however, Güntherodt's group in Basel obtained remarkable images of single vortices as well as vortex lattices in YBCO thin films by magnetic force microscopy [143, 144].

The magnetic force microscope—very similar to the atomic force microscope—consists of a tiny magnetic tip integrated on a cantilever of about 0.2 mm in length. This tip interacts with the stray magnetic field of the sample. The force on the tip is detected by measuring the deflection of the cantilever. During the experiment shown in Fig. 36 the tip has been scanning over a YBCO thin film at liquid nitrogen temperature (77 K); the individual vortices produce a repulsive force of less than 10^{-12} N. The number of vortices observed increases linearly with the applied external field, and one can easily show that individual vortices carrying one flux quantum ($\Phi_0 = 2.03 \times 10^{-15}\,\mathrm{V\,s}$) are imaged.

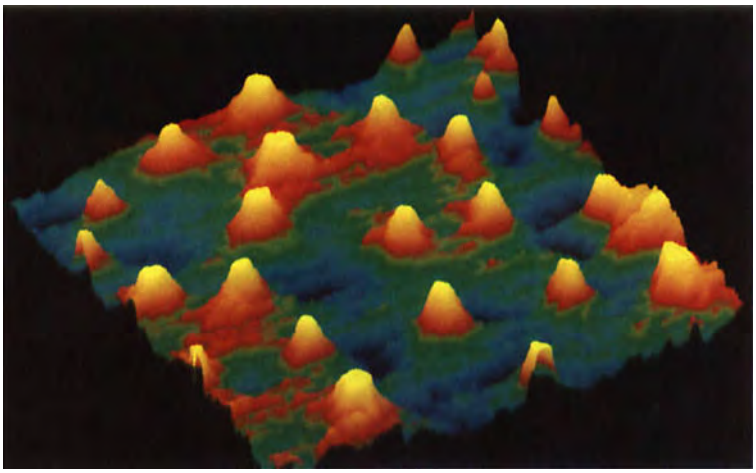


Figure 36. Magnetic force microscopy image of vortices in a YBCO thin film; the image shown is 22 × 22 μm in size. (Courtesy of A. Moser.)

8.2.6 Magneto-Optical Studies

Magnetic flux behavior and critical current density are key phenomena for the current-carrying properties of superconductors. Usually, measurements of magnetic properties give only integral or average data. A direct study of the magnetic flux structure by means of the magneto-optical technique (see Vol. I of this Handbook) provides us with local and much more detailed information. Recently, pushed by the increased interest in ceramic superconductors, this technique was further developed and successfully applied [145–150].

Using the high-resolution Faraday (HRF) effect technique, flux distributions in sintered superconducting samples can be obtained; it can be shown that the flux penetration occurs stepwise, and only the last step of the penetration of the Abrikosov vortices in the individual grains can be observed directly [147]. The high spatial resolution of the HRF technique enables the determination of the flux density profile inside individual grains. From these profiles the intragranular critical current densities and acting local pinning forces can be obtained. Figure 37 shows an SEM image of YBCO grains where BaO spheres (arrowed) can be detected at the grain boundaries. The flux distribution of such a polycrystalline YBCO sample is shown in Fig. 38; measurements were performed at 5 K, and at an external field of $\mu_0 H_{ext} = 273\,\mathrm{mT}$. The flux is found to penetrate the superconducting grains individually. The grain boundaries are clearly visible in the obtained flux distribution. Larger grains show the penetration of

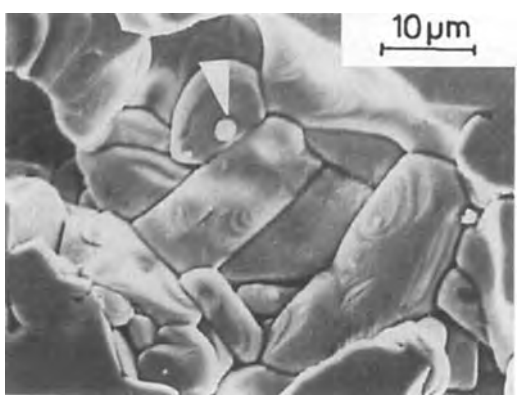


Figure 37. SEM image of $YBa_2Cu_3O_{7-\delta}$ grains. $BaCuO_2$ spheres at the grain boundaries can be detected (marked by a white arrow). Holes due to evaporated CO_2 are found in most of the grains. (Courtesy of Koblischka [144].)

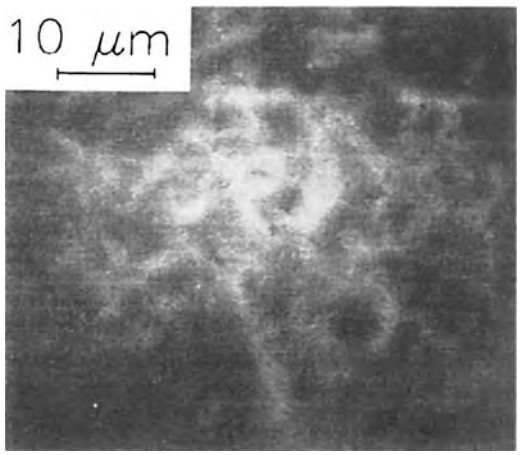


Figure 38. Flux distribution on a polycrystalline $YBa_2Cu_3O_{7-\delta}$ sample such as that of Fig. 36 at a temperature of 5 K and an external magnetic field of $\mu_0 H_{ext} = 273$ mT. The flux is found to penetrate the superconducting grains individually. Characteristic for granular samples, the grain boundaries are clearly visible in the flux distribution obtained. Larger grains also show a penetration of a flux front towards the center of the grain. The flux density gradients pointing towards the center of the grain are also clearly visible. (Courtesy of Koblischka [144].)

a flux front towards the center of the grain, indicating that each grain behaves similarly to a small single crystal in this field regime ($H_{ext} < H_{cl}$).

The magneto-optical technique can also be used to study the altered flux penetration in heavy ion-irradiated superconducting materials. Irradiations with high-energy metallic ions (e.g., 230 MeV nickel ions in [145]) is known to induce columnar tracks of amorphous material along the projectile paths (see later). These columnar tracks are found to be effective pinning centers if an external magnetic field is applied parallel to the tracks. Their radius of a few nanometers coincides approximately with the vortex core radius (for example, see Fig. 50 for lead irradiation in thallium-based material). These structural defects are well characterized by HREM. The HRF technique enables the obtention of direct local (spatial resolution 0.8 μm) information on the pinning behavior in superconductors. Observations in the excellent work of Schuster [145] were done at a temperature of 5 K; the external magnetic field was in all cases perpendicular to the sample surface and therefore parallel to the *c* axis of the material. The Shubnikov phase is imaged as bright domains; the Meisner phase remains dark. The left half of the crystal is the irradiated region, whereas the right half is unirradiated. Figure 39a shows the bright image of the crystal. In Fig. 39b–e the external magnetic field is raised from $\mu_0 H_{ext} = 171$ mT to 512 mT. In Fig. 39f the corresponding remanent state to Fig. 39e is shown; the domain of negative vortices and the annihilation zone separating the domains of different sign are clearly visible (marked by arrows). In this image the coupled behavior of the two kinds of vortice is shown. The negative flux clearly follows the penetration scheme of the positive flux, yielding the same domain pattern [145].

8.2.7 Less Conventional Imaging Techniques

Most of the less conventional methods for imaging high-T_c materials give very precise information, but for various reasons have not become routine techniques for the characterization of superconducting materials.

To investigate the structural properties of single crystals of the type Nd_2CuO_4, $(LaSr)_2CuO_4$, or $YBa_2Cu_3O_7$, investigators such as Bdikin et al. [151] are using

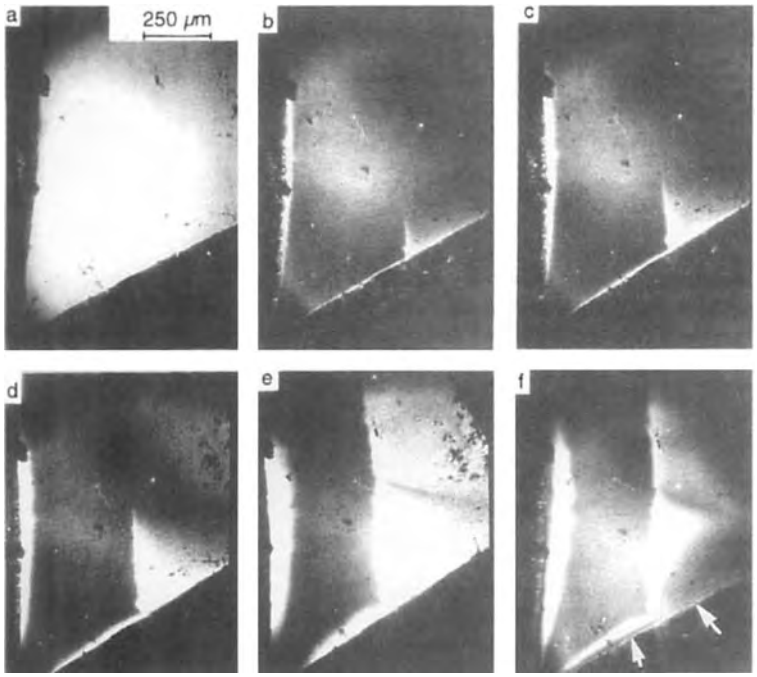


Figure 39. Domain pattern at 5 K of a $DyBa_2Cu_3O_{7-\delta}$ single crystal irradiated with 0.5 GeV ^{127}I ions. Only the left half of the crystal was exposed to the irradiation. (a) Bright field of the region, shown for comparison. (b) $\mu_0 H_{ext} = 171$ mT. (c) $\mu_0 H_{ext} = 256$ mT. (d) $\mu_0 H_{ext} = 341$ mT. (e) $\mu_0 H_{ext} = 512$ mT. (f) Corresponding remanent state $\mu_0 H_r = 512$ mT to (e); clearly visible are the domains of negative vortices as marked by arrows. (Courtesy of Schuster et al. [142].)

X-ray anomalous transmission effects (the Borrmann effect). The angular scanning topograms and Borrmann topograms allow them to deduce information about the presence or absence of an inversion center in the structure. The method clearly only holds for single crystals; for strongly distorted crystals the interpretation of the results is still problematic.

Surface information can be obtained not only by STM or AFM but also by other techniques such as scanning Auger microscopy (SAM) and angle-resolved photoelectron spectroscopy (ARPES). Most superconducting ceramics are highly anisotropic and have pronounced (001) cleavage planes, making them highly suitable for surface studies such as SAM or ARPES. For the bismuth superconductors the cleavage plane is clearly between the two BiO layers, but in YBCO there is a controversy on the terminating plane, being either between the BaO and the CuO_2 plane or between the CuO and the BaO layer. Different techniques (SAM [152], HREM [153, 154], STM [155, 156], ARPES [157, 158], and X-ray spectroscopy (XPS) [159, 160]) give different results, and often one technique provides contradictory answers. We therefore have to conclude that most probably YBCO cleaves on both planes. Schroeder et al. apply a special cleavage technique such that both surfaces from a single cleavage can be studied and correlated by SAM [161].

An important problem in high-T_c ceramics is the exact chemical composition: the deviation of the local composition with respect to the average composition or the nominal composition. Direct imaging techniques such as elemental mapping of the different element by energy dispersive spectroscopy (EDS), EELS, or Auger

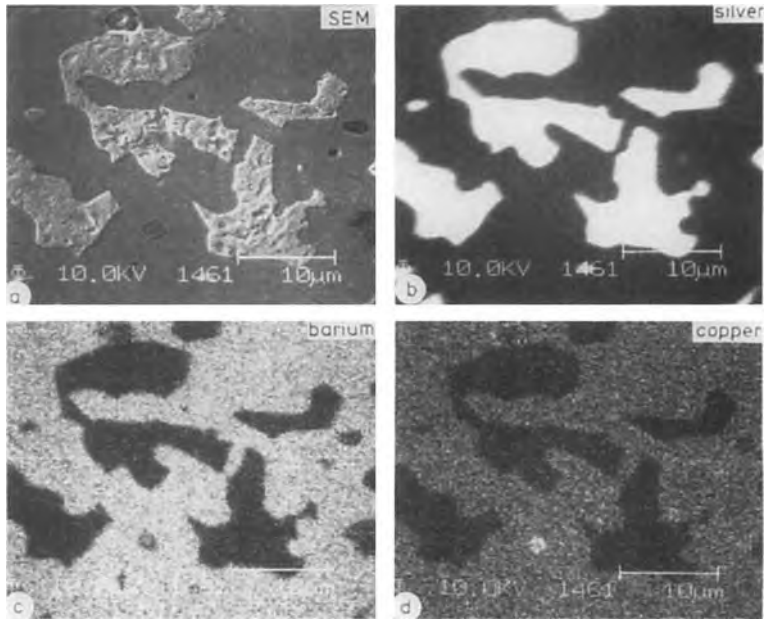


Figure 40. Typical $YBa_2Cu_3O_{7-\delta}$ and silver grains (SEM) of a $YBa_2Cu_3O_{7-\delta}$–30 wt.% silver sample after (a) polishing and various elemental AES maps of (b) silver, (c) barium, and (d) copper. (Courtesy of Udayan De [159].)

electron spectroscopy (AES) provide useful information for the synthesis of new superconducting materials. A good example is shown in Fig. 40 [162], where silver-added YBCO ceramic composites have been investigated by high-resolution scanning Auger electron microscopy. The silver particles within the matrix are clearly revealed, and the eventual effects of silver diffusion can be studied. The study of Udayan De [162] showed no detectable effects of interdiffusion of silver into the YBCO phase. A similar example comes from Mokhtari et al. [163], who studied the influence of iodine insertion on the YBCO superconducting properties by EDS and SEM.

As usual, the combination of different techniques is the most powerful. A beautiful illustration is given by Dravid et al. [164], who analysed the 'well-known' (001) planar defects in $YBa_2Cu_3O_{7-x}$, generated by decomposition of the material in a humid atmosphere. These defects were first detected by Zandbergen et al. [165] and correctly interpreted as double CuO layers forming locally the $YBa_2Cu_4O_8$ structure. Dravid et al. [164] brought direct chemical evidence by combining HREM observation of the defects together with local X-ray microanalysis and EELS of the defect under consideration. Figure 41 shows the HREM image and the X-ray and EELS spectra of the matrix and the planar defect. The EELS Cu-L_{23} edge from the fault region is compared with the Cu-L_{23} edges of the adjacent 1-2-3 matrix and with that of a CuO external standard. Visual inspection shows a remarkable similarity with the edges of the CuO precipitate. For the microanalysis a probe size of a little less than 1 nm was used. Although this is an example where results from the literature were only confirmed, it is clear that the method opens enormous possibilities for the analysis of

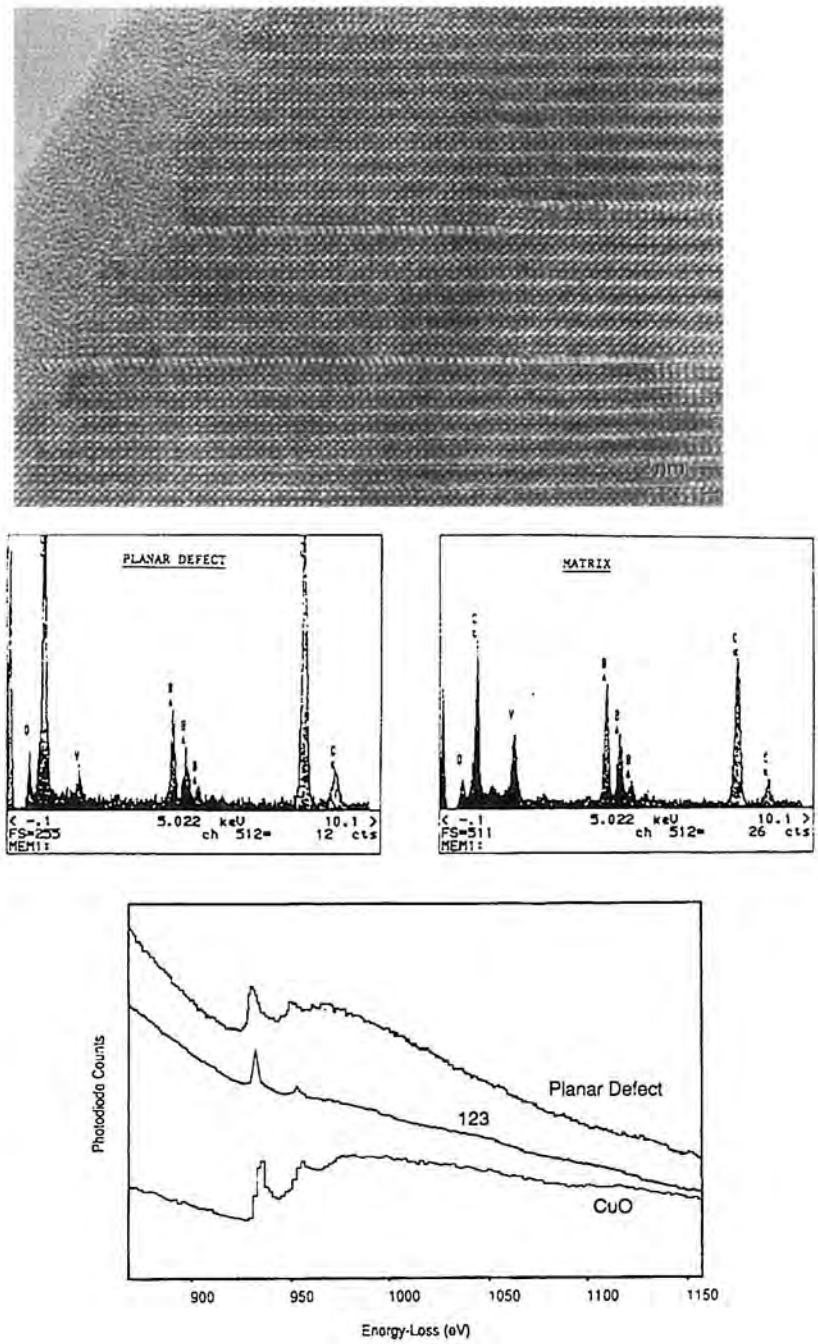


Figure 41. (a) HREM image showing a couple of (001) double layer planar defects in $YBa_2Cu_3O_{7-\delta}$. (b) X-ray spectra from the planar defect (left) with a <1 nm probe and from the matrix (right) with a larger probe. (c) EELS Cu-L_{23} edge from the fault region along with the Cu-L_{23} edges of the adjacent 1-2-3 matrix and a CuO external standard. (Courtesy of Dravid et al. [161].)

nanoprecipitates or nonstoichiometric defects within known superconductors. A similar combination of techniques is described by Faulques et al. [166] in their study of iodine insertion in high-T_c cuprates; they use Raman spectroscopy, XPS, and EELS to locate the iodine in bismuth, thallium, and YBCO superconductors.

Although field emission microscopy or spectroscopy was the first technique to provide atomic resolution (for an excellent overview see [167]) and although it does provide information about the electronic properties of the sample surface, it is not a very popular technique for the study of superconducting materials. It has been used by Shkuratov et al. [168] to study the evolution of the Fermi level at various temperatures between room temperature and about 100 K. The fact that atomic resolution is more easier attained by other techniques, and the stringent conditions on sample preparation, has clearly retarded the applications of field emission microscopy techniques.

Methods to investigate the spatial variation of the critical temperature T_c and the critical current J_c provide relevant information for the development of high-T_c superconducting electronic devices. When the devices contain insulating layers, SEM can no longer be used because of charging effects. Low-temperature scanning laser microscopy (LTSLM), however, can be used for this purpose. Details of the experimental set-up can be found in the literature [169, 170]. The method not only provides the local variations in T_c or the weak links with reduced critical current but is also able to characterize every layer of a multilayer device separately [170].

Also, the imaging of local electrical inhomogeneities in thin films at room temperature is an important feature, since they do not only reduce the reproducibility of the main parameters, but may also result in an excess of the $1/f$ noise in the junctions. By laser irradiating the thin films and monitoring the spatial distribution of the laser beam-induced electrical response in a thin film, one is able to image electrical inhomogeneities such as microbridges or grain boundary junctions [171].

Flux creep measurements are normally performed by SQUID (superconducting quantum interference device) or torque magnetometers, but flux creep effects can also be directly imaged [172]. They use the property of the high-resolution Faraday effect to allow dynamic observations of the Shubnikov phase and directly observed effects of flux creep.

8.3 Specific Problems in Superconductivity Research

8.3.1 Surface Studies by Transmission Electron Microscopy

Although SEM, STM, and AFM or even AES, XPS, or low-energy electron diffraction are more straightforward techniques to provide information on the surface, TEM can provide important information, inaccessible by other techniques. HREM has the advantage of providing at the same time an image of the surface as well as of the underlying structure. For the $YBa_2CU_3O_7$ structure—as for all superconducting ceramics based on the perovskite structure—the cleavage plane is the (001) plane, but the terminating plane is found to be the BaO plane or the

CuO plane, never the superconducting CuO_2 plane [173, 174, 175]. On the other hand, a very detailed and careful study has been made by Xin et al. [176]; they found the terminating layer to be the CuO_2 plane, corresponding to a sequence of bulk–Y–CuO_2. The results are confirmed by different image simulations where the different possible models are compared with experiments. The same authors also suggest partial oxygen loss at the terminating plane; the evidence for this, however, is less convincing. For the very anisotropic bismuth-based superconductors the preferential cleavage plane is clearly between the two BiO layers [177]. For the more exotic compound

$$Pb_2Sr_2Ca_{0.5}Y_{0.5}Cu_3O_{8+\delta}$$

—abbreviated to PSCYCO—which has a stacking

$$-CuO_2-(Ca, Y)-CuO_2-SrO-PbO-$$
$$(Ca,Y)-PbO-SrO-$$

the surface plane is the SrO layer, indicating that the coupling between the SrO layer and the PbO layer is the weakest; this can be deduced from the high-resolution image of Fig. 42 [178] by comparing the experimental image with the calculated images.

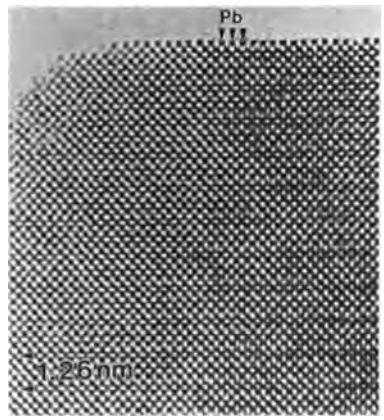


Figure 42. [110] HREM image of PSYCO, showing a (001) surface. In the top layer the terminating black dots can be identified as lead atoms.

8.3.2 Characterization of Superconducting Thin Films

Most of the recently discovered superconducting compounds can be grown in thin film form. This growth proceeds mostly by sputtering, laser ablation, chemical vapour deposition, or molecular beam epitaxy on substrates of different types and different quality. Most attention by far has been devoted to the growth of $YBa_2Cu_3O_7$ on different substrates. Most superconducting thin films have been grown on MgO or $SrTiO_3$; however, more exotic substrates such as $LaAlO_3$, $NdGaO_3$, or Mg_2TiO_4 are also used [179–199]. More recently, films have been successfully grown on technologically relevant single-crystal substrates such as silicon, GaAs, Al_2O_3, or SiO_2. Silicon and GaAs have a large misfit with respect to YBCO, and tend to react at the growth temperature, inducing a strong chemical interaction and a rough surface with microcracks [183]. These problems can be overcome by the introduction of buffer layers, intended to act as diffusion barriers and to enable the relaxation of thermal stresses. Buffer layers reported in the literature include yttria-stabilized zirconia (YSZ), Y_2O_3, CaF_2, SrF_2, BaF_2, or the semiconducting $PrBa_2Cu_3O_7$ [184]. Recently, complex barrier layers have been used in order to optimize the epitaxy and the physical parameters. YBCO films on silicon now have critical temperatures around 90 K and critical currents over

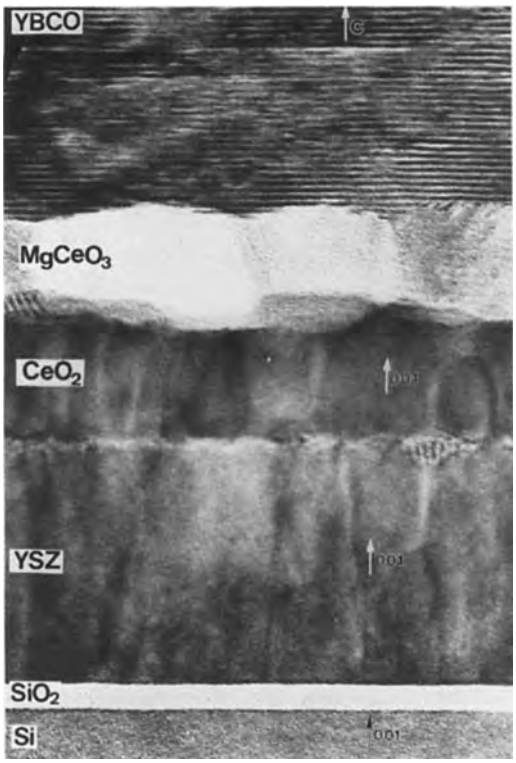


Figure 43. HREM image of a YBCO film on an MgO–CeO_2–YSZ buffer.

10^6 A cm^{-2} at liquid N_2 temperature [185]. An example of such a film with an extremely complex buffer sequence is shown in Fig. 43. The direct growth of Y_2O_3 or CeO_2 on silicon leads to polycrystalline film formation where grain boundaries are assumed to be diffusion paths and the surface of these films is found to be corrugated. However, a single-crystal oriented YSZ buffer can be formed on silicon in spite of the presence of a 3–5 nm thick amorphous layer on the YSZ–Si interface. For microstructural applications, including Josephson-based and high-frequency devices, a specific YBCO film orientation is required. For *c* axis-oriented YBCO films, different in-plane rotations can be achieved by sandwiching additional MgO seed layers. It is seen, however (Fig. 43), that there is a strong interdiffusion between the CeO_2 layer and the MgO layer, forming the $MgCeO_3$ compound [185].

A lot of attention is being devoted to the optimization of such sequences of thin films. As well as after the deposition, imaging techniques play a major role in the characterization of the films and the film quality during deposition. The most useful techniques for this purpose are SEM, AFM, STM, and TEM. Very often the combination of different techniques provides the most complete information; examples of these are given in the literature [186–191].

The surface morphology and roughness, extremely important in thin film technology, is best studied by STM or AFM. A beautiful illustration is given by Lang et al. [191] in their study of the different growth stages of *c* axis-oriented $YBa_2Cu_3O_7$ thin films on $LaAlO_3$. For a reliable study it is desirable to have films with different thicknesses prepared under similar deposition conditions using identical targets and substrate materials. To obtain this, Lang et al. use the 'half shadow technique' [191] whereby thin films with a thickness gradient are obtained. This allows them to study by STM the nucleation and growth of two-dimensional islands (Fig. 44), their coalescence, and the formation of well-developed growth hills (Fig. 44) and a continuous film. The growth mechanism is clearly a layer growth with respect to the height of the YBCO unit cell. In addition to the normal island growth, spiral growth has also been observed by STM (Fig. 45). Gao et al. [188] observed in this way a 'brick wall' microstructure during the growth of a YBCO film on an MgO

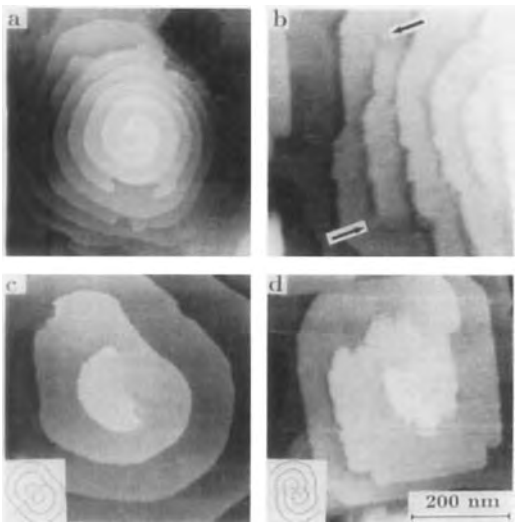


Figure 44. Different kinds of growth hills with screw dislocations: (a) single right-handed screw dislocation with about 8 turns of growth steps; (b) annihilated pair of dislocations (marked by arrows) of opposite sign; (c) a pair of screw dislocations of opposite sign further apart from each other than 2rc, sending out loops of steps; and (d) pair of screw dislocations of like sign with a step activity indistinguishable from that of a single spiral. The scale bar is valid for all figures. (Courtesy of Lang et al. [188].)

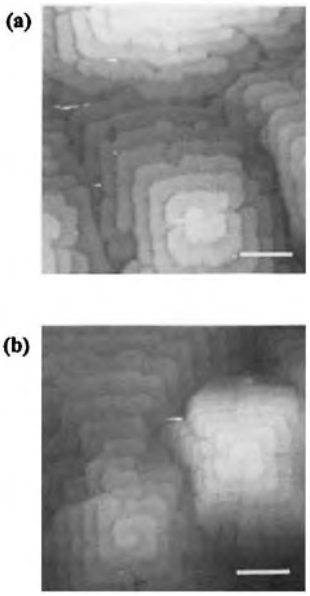


Figure 45. STM images (the white markers correspond to a length of 100 nm) reveal the surface topography of a 100 nm thick $YBa_2CU_3O_7$ film obtained by 90° off-axis magnetron sputtering. The presence of terraces directly results from the epitaxial, layer by layer growth process. The *c* axis of the film is perpendicular to the (001)-oriented MgO substrate. The film is composed of islands growing along screw dislocations. The screw dislocations are anticlockwise in (a) and clockwise in (b). (Courtesy of C. Van Haesendonck.)

substrate. This apparently results from the overlap between terraces belonging to adjacent spiral-shaped islands.

8.3.2.1 Thin Film Sample Preparation for Transmission Electron Microscopy

It is clear from the foregoing discussion that in the study of thin films (not only superconducting thin films) HREM plays an extremely important role. Preparing reliable and reproducible samples for electron microscopy, without altering or destroying the film or the substrate–film interface is therefore crucial. A lot of attention has to be devoted to produce so called 'cross-section' samples, where both the film and the substrate can be revealed over distances as large as possible. The initial sample is generally of the order of 10 mm × 10 mm × 500 μm, the last being the thickness of the support. Such a sample is sliced (a few hundred micrometre widths) by a diamond saw or wire saw perpendicular to the film. Then the different slices are glued together, generally with the superconducting thin films facing each other; sometimes it is advisable to glue them all in the same direction [193] (Fig. 46). To reduce the ion milling time, the glued stack is mounted on a tripod holder [194] and thinned down to 10 μm; details are given by Traeholt et al. [193, 196]. During ion milling, the incident angle is kept as small as possible, and one

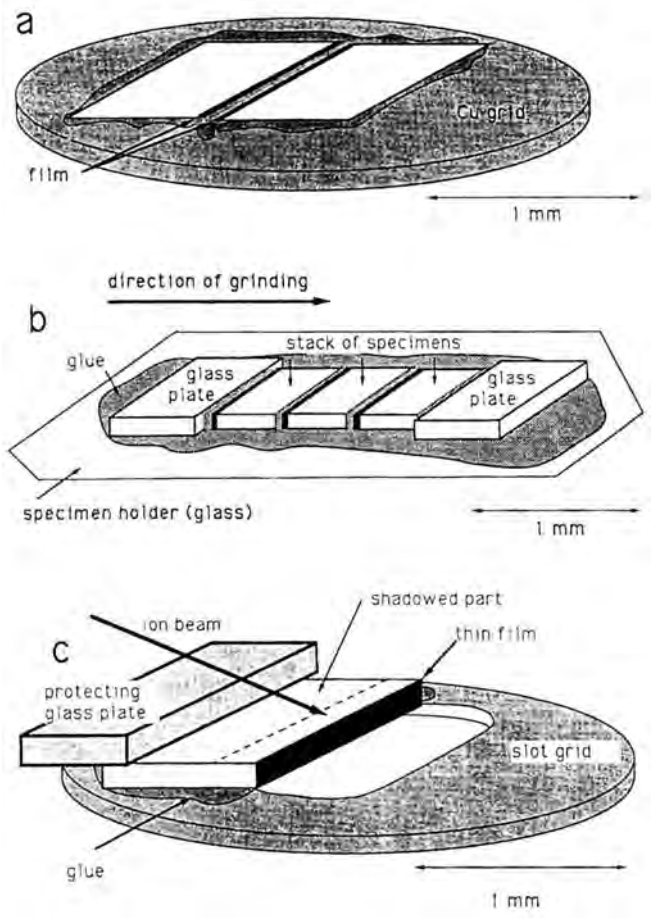


Figure 46. (a) The most common way to make TEM cross-sections. The cross-section is mounted on a support grid with the film sides facing each other. Ion milling of the cross-section is done with two guns while the cross-section is rotating. (b) All film sides are facing the same direction. The grinding direction is opposite to the normal of the film as indicated. The glass pieces function as protection. (c) The set-up for ion milling. A major part of the cross-section is shadowed by the glass piece to keep a thick supporting region. The ion milling is done with one gun without rotation of the specimen. (Courtesy of Traehold et al. [190].)

has to verify regularly by OM the thickness of the sample. Traeholt describes a special shadowing method using a glass piece, a nonrotating sample set-up, and ion milling from one side only [196]. With the recent ion milling machines, using a retarding field, incident angles down to 1° can be obtained, producing large electron-transparent areas.

8.3.2.2 Precipitate Identification

Another important feature in the thin film production is the presence—or the absence—of foreign precipitates. Some of these precipitates (e.g., CuO, BaCuO, $YCuO_2$, Y_2BaCuO_{5-x}, or $Y_2Cu_2O_5$, originating from unreacted material) are to be avoided; others such as nanoprecipitates of Y_2O_3 could have a favorable effect on the flux pinning properties of the material. Virtually the only technique to study these precipitates and to determine their structural relationship with the YBCO is the combination of HREM, ECX, and nanodiffraction. Innovative work in this field has been done by Eibl and co-workers [179]. The small size of these precipitates has provoked some ambiguity about the

composition and the orientation relationship with the underlying substrate. In a recent paper by Verbist et al. [192] the different proposals are tested with respect to each other; two different epitaxial relationships are found. The YBCO *c* axis can be parallel to the [001] or the [110] axis of Y_2O_3. The [110] precipitates are only situated near the foil surface. The information from HREM, together with the Fourier transform and the simulated images, allow identification of the precipitate unambiguously as Y_2O_3. The epitaxial relationship can be deduced from the HREM or from a nanodiffraction experiment.

8.3.2.3 Substrate–Superconducting Film Interface

In order to grow epitaxial films and to have a better understanding of the interface problems it is important to have a detailed image of the substrate–film interface or the interface between subsequent buffer layers. Cross-section electron microscopy is virtually the only technique providing reliable and detailed information. Innovative work in this field has been done by Eibl and co-workers [179, 195] and by the group around Zandbergen [180, 193, 196, 199]. The latter have studied convincingly the YBCO–MgO interface of a thin film produced by laser ablation. Concerning the first atomic layer on the MgO substrate, a number of ideas were put forward. Tietz et al. [197] suggested that the first YBCO layer is the—superconducting—CuO_2 layer; this would indeed provide the best continuation of the oxygen sublattice. Pennycook et al. [198] reported that the thermodynamically lowest energy is obtained for the interface with the CuO chain plane as the first layer. They performed *Z*-contrast on laser-ablated thin films and found that the first layer is indeed the CuO layer or the CuO_2 layer next to the BaO layer. Traeholt et al. [196] showed by HREM that the first layer is indeed the CuO layer but that two types of interface configurations occur:

(1) The YBCO lattice and the MgO lattice continue up to the interface; this often creates a periodic strain in the substrate.
(2) The YBCO lattice and the MgO lattice are separated by an amorphous layer with a thickness of the order of the two atomic layers; in this case most strain is released.

A more complex example of growth of YBCO on a YSZ substrate is illustrated in Fig. 47 (from the work of Wen et al. [199]). In all samples they observed the formation of an intermediate layer of $BaZrO_3$, located between the YSZ substrate and the superconducting film. HREM allowed the determination of the stacking sequence of the $BaZrO_3$–YBCO interface as $(BaZrO_3)$–ZrO_2–BaO/CuO–BaO–(YBCO) with the CuO layer as the initial YBCO layer. Occasionally the BaO layer was also found to be the initial layer.

8.3.2.4 Towards Applications

High-T_c $YBa_2Cu_3O_{7-x}$ films can be used in hybrid circuits integrating active superconducting elements and interconnections with semiconductor devices. The chemical interaction between YBCO and silicon, the relatively large mismatch in thermal expansion coefficients, and the lattice parameters hamper the use of silicon in

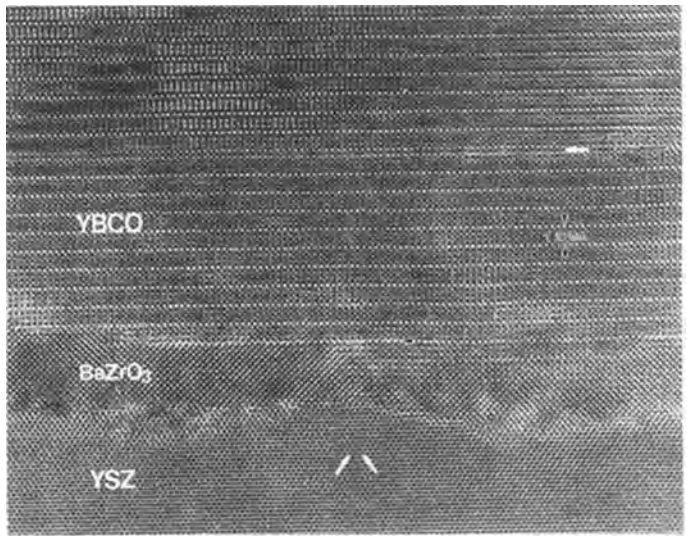


Figure 47. HREM image of a YBCO film on a YSZ substrate taken along [110] of YSZ. A reaction layer of $BaZrO_3$ between YSZ and YBCO is observed. Steps at the $BaZrO_3$–YBCO interface correspond to a whole unit cell of YBCO. The black arrowhead indicates that the CuO layer as the initial layer of the YBCO film. The white arrow indicates a stacking fault. The interfacial dislocations are indicated by two white arrows. (Courtesy of Wen et al. [196].)

direct combination with YBCO. Different buffer layers are used to eliminate the chemical interaction [200–204]. High-quality YBCO films can be grown on double CeO_2–YSZ buffer layers. An additional MgO seed layer on top of the CeO_2–YSZ buffer allows the introduction of in-plane rotations of the YBCO film. This effect can be used in bi-epitaxial technology for the fabrication of Josephson junctions (JJs) where differently oriented parts of the YBCO film form controllable, isolated grain boundaries which are suitable as weak links [205, 206].

An optical micrograph of three microbridges is presented in the Fig. 48a, and an enlarged image of the central microbridge is seen in Fig. 48b. In the center of the bridge the line which separates two parts of the substrate with and without seed layer is visible.

High- and low-angle tilt grain boundaries in the YBCO film on an MgO–CeO_2–YSZ buffer can be analysed by electron microscopy. At low magnification the high-angle grain boundaries appear to be wavy; however, careful inspection by HREM reveals that the high-angle grain boundaries are mostly faceted. An area of distorted contrast at the grain boundary is confined within one or two YBCO unit cells from each side. The grain boundary shown in Fig. 49a also shows a darker band near the interface; in that area the lattice image shows more pronounced {110} fringes. This is more clearly illustrated in the enlargement of Fig. 49b. Simulated images under different conditions show a correlation between this lattice distortion and the local oxygen

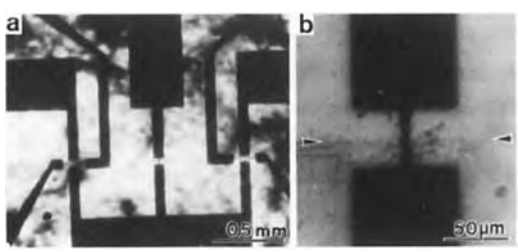


Figure 48. Optical microscope image of patterned YBCO film on a buffered silicon substrate. (a) Three microbridges in parallel. (b) Enlarged part of the central microbridge; the arrows show the boundary between parts with and without MgO seed layer.

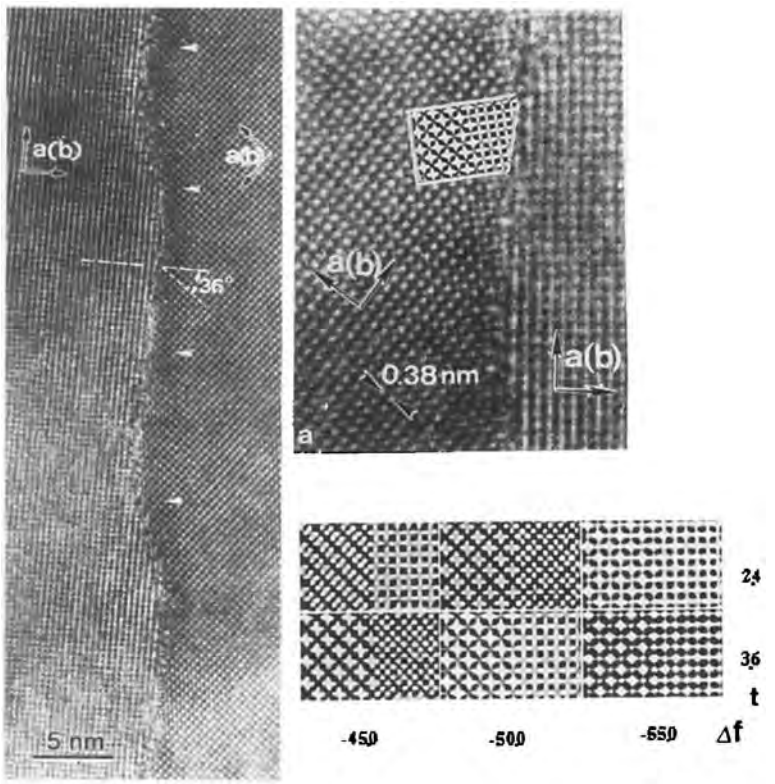


Figure 49. Images of an asymmetric grain boundary with the habit plane close to {100} or {010}. (a) A 36° grain boundary; note the dark band along the grain boundary (marked by white arrows). (b) Enlarged part of the grain boundary with an inserted simulated image based on a local oxygen depletion.

content. In a very similar way, extended defects occurring at the grain boundary can be analysed [207].

8.3.3 Ion Irradiation of Superconductors

In order to investigate the interaction of crystal defects with flux lines, several groups have introduced defects of a known size and number into well-characterized material. High-energy heavy ions produce latent tracks of several nanometers in diameter along their path through crystalline matter. These defect lines are known to act as pinning centers. Gerhäuser and co-workers studied the effect of 0.5 GeV ^{127}I irradiation in the bismuth 2212 superconductor [208, 209]. Zhou et al. [210] studied the microstructure of the same material after 12 MeV iron ion bombardment. Watanabe et al. [211] used 173 MeV xenon irradiation in YBCO films. Irradiation of thallium 2223 superconducting materials with Xe ions were performed by Hardy et al. [212]. The results for the different heavy ions and different materials are very similar: when observing along the ion irradiation direction, discs with a diameter around 5–10 nm are visible (Fig. 50a) with a density corresponding to the applied dose. Tilting experiments or stereo microscopy reveal the three-dimensional cylindrical shape of the heavy ion tracks. The structure of the damaged regions is revealed by HREM

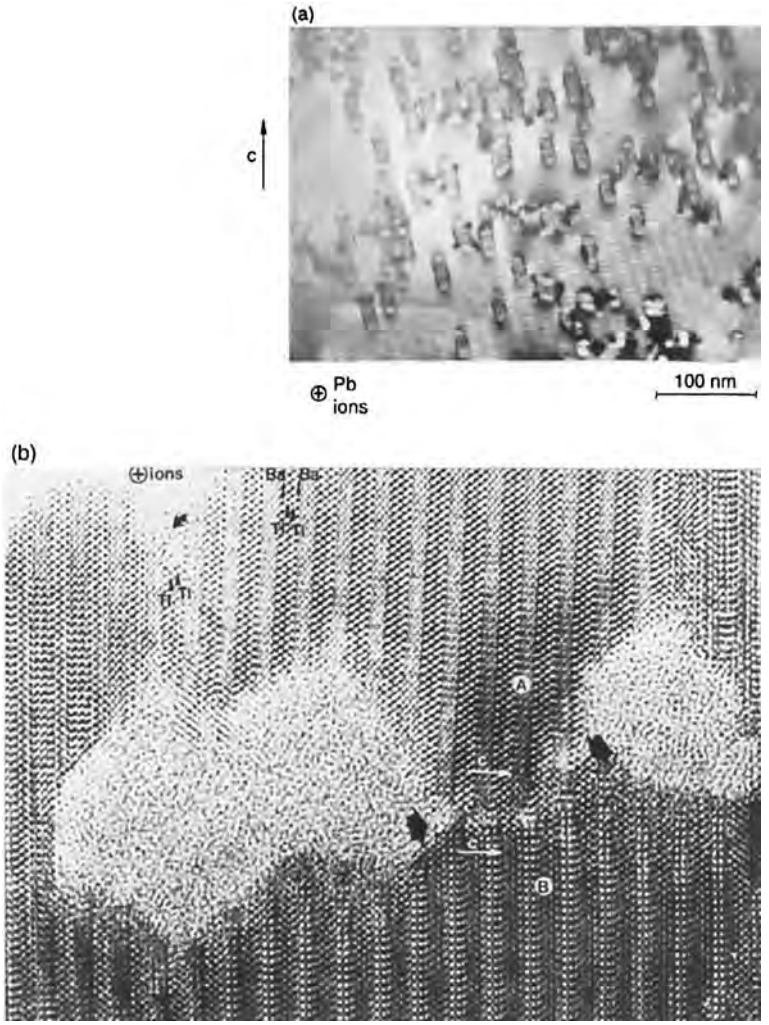


Figure 50. (a) Thallium 2223, lead-irradiated material, recorded with the beam almost parallel to the ion path. (b) [010] HREM image of lead-irradiated thallium 2223; the nature of the ions is reproduced in the top part of the image. The curved arrows indicate a microcrack at the level of the TlO layers.

(Fig. 50b); the inner part of the track is amorphous, and in the outer part some reconstruction has often occurred (Fig. 50) [213]. In a recent experiment we have irradiated different thallium superconductors (2212, 1223, and 2223). Ceramic plates of 150 μm thickness were irradiated at 300 K with lead or uranium ions of 6 GeV; total fluences varied between 5×10^{10} and 2×10^{11} ions cm^{-2}. The average diameter of the ion tracks is 9–10 nm, irrespective of the ion used and the fluence. A strain field is clearly visible around the tracks. The nature of the tracks can be determined by HREM: the core of the tube is clearly amorphous. Some of the projected traces are not circular (Fig. 50); this is attributed to a slow recrystallization which is sometimes observed 'in situ' under the electron microscope. As a result of the high strain, specific defects are formed, as shown in Fig. 50b, which is a [010] HREM image where the positions of the atoms are imaged as bright dots. Three

tracks are evident; the recombination of the structure at the interface between the amorphous and the crystalline areas can be observed. In between two neighbouring tracks, microcracks are formed in the crystal, and parts A and B are rotated over an angle of about 2° with respect to each other. The crystal apparently cracks at the position of the TlO layers (see curved arrow in the left part of the image). The formation of such defects is of primary importance to understand the physical behavior of irradiated materials.

Proton irradiation (e.g., see Kramer et al. [214]) at energies of 200 MeV gives completely different results. Immediately after irradiation, no defects were observed, but after ageing for over 90 days, the same samples showed defects with an anisotropic strain contrast. TEM shows that two types of defects are formed, small clusters (0.5 nm) with no special orientation and extended defects between 2 and 7 nm in length and 0.2 nm in thickness which concentrate along {100}⟨010⟩. The explanation given by Kramer et al. [214] is that protons predominantly produce point defects of copper and oxygen, which cluster over time. The effect of neutron irradiation on high-T_c materials has been well summarized and reviewed by Kirk and Weber [215]. Upon irradiating YBCO one observes a dark, strain-induced contrast of circular defects with a mean size of 4.5 nm at a fluence of $3 \times 10^{23}\,m^{-2}$ ($E > 0.1$ MeV) [216]. The defects, which are shown to be responsible for the increased pinning, can be studied by TEM. The defects, with a size of about 2–3 nm, are most probably amorphous or heavily disordered structures [217]. Zandbergen et al. [218] reported on the limitations on the monitoring of neutron-induced defects by TEM. Their samples were either ion milled or crushed and deposited on holey carbon film. Zandbergen et al. claim that the samples deteriorate inside the electron microscope due to an intercalation, leading to the appearance of planar defects; Frischherz et al. [217], however, show that by taking enough care during the experiment, the electron microscopy results are perfectly reliable.

Using TEM one can also follow the irradiation in situ [219]. Room temperature irradiation of YBCO-type single crystals with 200 keV neon ions or 300 keV xenon ions have been performed inside the electron microscope. The defects produced are (i) mobile extended defects which recover rapidly under normal observation conditions and (ii) amorphous areas which are stable and can be observed by HREM (Fig. 51).

8.3.4 Grain Boundary Structures

Establishing the structure–property relationship for grain boundaries has been the

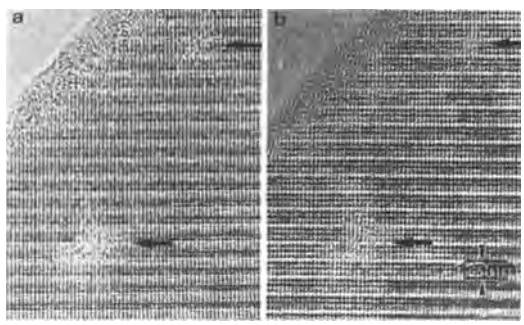


Figure 51. [100] HREM images of $YBa_2CU_3O_7$ irradiated with 300 keV xenon ions inside the electron microscope; amorphous radiated areas of about 1 nm in size are clearly visible.

goal for materials scientists and physicists for many decades. Although a clear relationship exists, the details of this relationship remain elusive. Critical current transport across grain boundaries of oxide superconductors is such an example [220]. Grain boundaries can be imaged by HREM [221–227], often by imaging along an axis common to both grains (e.g., see Fig. 49). This technique allows the deduction of the atomic configuration in the vicinity of the interface, but provides little or no information on the chemical composition in the immediate vicinity (first few nanometers) around the grain boundary. Detailed comparison with computer-simulated images partially overcomes this problem, but one has to be extremely careful with the obtained information, because slight thickness, focus, or orientation changes will considerably alter the experimental image [185].

Better compositional results can be obtained by combining HREM (structural information) with other techniques, such as EDX or EELS, better suited to obtain the local composition. Dravid et al. [220] probed in this way the hole concentration along the grain boundary in the 1-2-3 YBCO structure; they note significant changes in the O_K pre-edge peak by EELS. The changes are correlated to the hole concentration changes in the p-type cuprate superconductor. Similar work has been done by Alexander et al. [228], where grain boundaries in YBCO are imaged and structurally analysed by TEM, but where complementary compositional information is obtained from analytical electron microscopy and Auger electron spectroscopy. These authors systematically observe an excess of copper and a deficiency of oxygen at the boundaries.

A special and very promising technique to characterize materials in general and grain boundaries in particular is the so called *Z*-contrast technique (see Vol. 1, Chap. 4, Sec. 11.11). Using a combination of high-resolution *Z*-contrast imaging and EELS in a scanning transmission electron microscope, the local hole depletion and the local atomic structure can be studied simultaneously. The atomic resolution *Z*-contrast can be used to position the probe for energy loss spectroscopy, which by the analysis of the pre-edge feature of the oxygen K-absorption can be correlated with the hole depletion to an accuracy of 5% [229]. They conclude that no significant hole depletion has taken place in the symmetric boundaries, whereas in the asymmetric boundaries they measure a significant depletion in a zone extending over a far greater area than the disordered structure in the image. An example of the *Z*-type contrast is shown in Fig. 52.

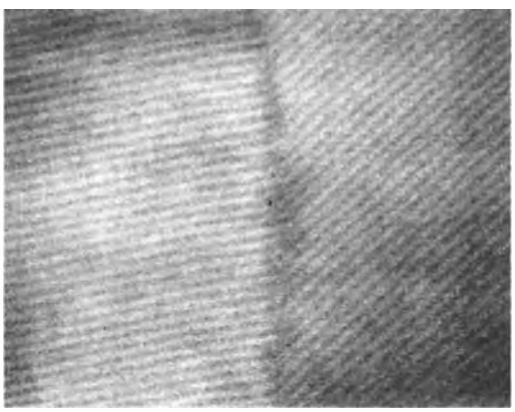


Figure 52. *Z*-contrast image of the asymmetric 29° (near Σ17 (001) 28.7°) grain boundary. The image shows a clear disordered/amorphous region at the grain boundary. (Courtesy of Browning et al. [226].)

8.3.5 Synthesis and Development of New Superconducting Phases

Numerous investigations have been performed in copper oxides since the discovery of superconductivity. The structure of layered cuprates show great flexibility to accommodate various cations in the normal layers which ensure the connection between the superconducting copper layers. The role of these noncopper cations in superconductivity is much more important than was expected at the beginning of the search for high-T_c superconductors.

Electron microscopy has played an extremely important role in the development of new superconducting materials. In some samples only traces of superconductivity are found, but when the phase responsible for the superconducting behavior can be identified and its composition determined, a new superconducting material is born. In other cases the presence of planar defects in existing superconducting materials provides the clue for the development of new superconducting materials. An extremely good example of this is the discovery of double (CuO) layers in normal 1-2-3 YBCO material [50], which has led to the development of the 1-2-3.5 [53] and 1-2-4 [54] superconductors. Further experiments have led to structures such as the 1-2-5 [230–232] or even the 1-2-6 [233] phase; however, these phases are only present in very local areas or are formed under the electron beam of the microscope.

Electron microscopy has also played an important role [234–238] in the development of the infinite-layer superconductors such as $Ca_{1-x}Sr_xCuO_2$ [231] and related structures. A large number of bismuth-, thallium- and mercury-based superconducting compounds of the type '1 2*n* *n* + 1' and/or '2 2*n* *n* + 1' [239–246], and very often TEM imaging was the key for their development. However, particularly with the recent development of superconducting oxycarbonates and mercury-based superconductors, TEM has played an extremely important role. I will illustrate this with some of our recent work, a collaboration between the University of Caen, the CNRS Grenoble, and the University of Antwerp.

Development of Superconducting Oxycarbonates

The '1-2-3' derivatives

$$(Y,Ca)_n(Ba,Sr)_{2n}Cu_{3n-1}CO_3O_{7n-3}$$

whose structure consists of an ordered replacement of rows of CuO_4 groups by CO_3 groups in $YBa_2Cu_3O_7$, are the first example of superconducting oxycarbonates. Two oxides of this series have been shown to exhibit superconducting properties — $(Y_{1-x}Ca_x)_{0.95}Sr_{2.05}Cu_{2.4}(CO_3)_{0.6}O_y$ [247] and $YCaBa_4Cu_5(NO_3)_{0.3}(CO_3)_{0.7}O_{11}$ [248]; both correspond to the members $n = 2$ of the above series. These compounds have a T_c (onset) of 66 and 82 K, respectively, with superconducting volume fractions of 5 and 15%, respectively, at 5 K.

The second type of oxycarbonates to exhibit superconducting properties is represented by the phase

$$(Ba_{1-x}Sr_x)_2Cu_{1-y}O_{2y+x}(CO_3)_{1-y}$$

[249, 250] with a T_c (onset) of 40 K and a Meissner volume fraction of 20%. Its structure belongs to the $Sr_2CuO_2CO_3$ type, that is, it consists of octahedral

copper layers connected through rows of carbonate groups. The close relationship between the $Sr_2CuO_2CO_3$-type structure and the perovskite structure suggests that the CO_3 groups should also be able to accommodate all the layered cuprates, that is, layered structures involving rock salt and perovskite layers. This is the case for the recently discovered superconductors $Tl_{0.5}Pb_{0.5}Sr_4Cu_2CO_3O_7$ [251], $Tl_{1-x}Bi_xSr_4Cu_2CO_3O_7$ [252], $Bi_2Sr_4Cu_2CO_3O_8$ [253], and $Bi_2Sr_{6-x}Cu_3O_{10}(CO_3)O_{10}$ [254].

The oxycarbonate

$$Tl_{0.5}Pb_{0.5}Sr_4Cu_2CO_3O_7$$

exhibits a critical temperature up to 70 K with a Meissner volume fraction of 70% at 5 K. The structure of this tetragonal phase consists of the intergrowth of double rock salt layers $[(Tl_{0.5}Pb_{0.5}O)(SrO)]_x$ and single perovskite layers, linked through layers of CO_3 groups. The oxycarbonate $Tl_{1-x}Bi_xSr_4Cu_2CO_3O_7$, which exhibits a similar structure, is characterized by a lower critical temperature ($T_c = 54$ K) but also a large Meissner volume fraction of about 60%. In fact, the structure of these two superconductors can be described as an intergrowth of the two structures $Sr_2CuO_2CO_3$ and '1201', that is, $Tl_{0.5}A_{0.5}Sr_2CuO_5$ (A = Pb or Bi). As a consequence, these oxides can be considered as the first members ($n = n' = 1$) of the intergrowth with the general formula $(Tl_{1-x}A_xSr_2CuO_5)_n(Sr_2CuO_2CO_3)_n$. The remarkable feature of these two phases deals with the fact that the two structures which form their framework either do not superconduct or exhibit a lower critical temperature. $Sr_2CuO_2CO_3$ has indeed never been observed to superconduct, whereas only traces of superconductivity (less than 1%) up to 60 K were detected in the 1201 phase $Tl_{0.5}Pb_{0.5}Sr_2CuO_5$ [252], which might be due to the presence of traces of this oxycarbonate. In a similar way, the 1201 phase $Tl_{1-x}Bi_xSr_2CuO_5$ was found to exhibit a significantly lower T_c of 45 K (onset) with a broad transition [255, 256]. Thus, it clearly appears that the intergrowth of the two nonsuperconducting structures, the oxycarbonate $Sr_2CuO_2CO_3$ and the 1201 phase, significantly improves the superconducting properties.

In the same way, the two superconductors $Bi_2Sr_4Cu_2CO_3O_8$ [253] and $Bi_2Sr_{6-x}Cu_3(CO_3)_2O_{10}$ [254] correspond to the members $n = n' = 1$ and $n = 1$, $n' = 2$ of the intergrowth

$$(Bi_2Sr_2CuO_6)_n(Sr_2CuO_2CO_3)_{n'}.$$

Figure 53 shows, as an example, the schematic structure of $Bi_2Sr_4Cu_2CO_3O_8$,

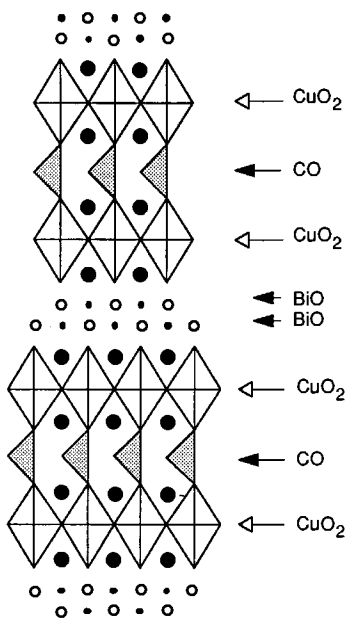


Figure 53. Schematic structure of the 30 K superconductor $Bi_2Sr_4Cu_2CO_3O_8$, the $n = n' = 1$ member of the intergrowth series $(Bi_2Sr_2CuO_6)_n(Sr_2CuO_2CO_3)_{n'}$.

which consists of an intergrowth of single 2201 layers with single $Sr_2CuO_2CO_3$-type layers. The second member, $Bi_2Sr_{6-x}Cu_3(CO_3)_2O_{10}$ is deduced from the latter by replacing the single $Sr_2CuO_2CO_3$-type layers by double oxycarbonate layers.

The most striking feature is the value of the critical temperature: 30 K for $Bi_2Sr_4Cu_2CO_3O_8$ and 40 K for $Bi_2Sr_{6-x}Cu_3(CO_3)_2O_{10}$; both T_c values are higher than those of the mother structures. $Bi_2Sr_2CuO_6$ exhibits a maximum T_c of 22 K [257], whereas $Sr_2CuO_2CO_3$ does not superconduct. Also spectacular is the increase of the T_c from 30 to 40 K as the thickness of the nonsuperconducting layer $(Sr_2CuO_2CO_3)_{n'}$ increases.

The recent discovery of the 62 K superconductor $TlSr_{4-x}Ba_xCu_2CO_3O_7$ [258] demonstrates the great ability of CO_3 groups to occupy cationic sites different from the copper sites. This phase indeed exhibits a very original structure (Fig. 54) closely related to that of $Tl_{0.5}Pb_{0.5}Sr_4Cu_2CO_3O_7$; and was determined and analysed by HREM. The structure of $TlSr_2Ba_2Cu_2CO_3O_7$ can be obtained by a shearing mechanism along $\bar{c}$, that is, by a translation over $\bar{c}/2$ of the latter structure every four CuO_6 octahedra. From this shearing mechanism, it appears that the $[CuO_2]_x$ layers of each block remain unchanged and form infinite layers parallel to (001). This is not the case for the $[TlO]_x$ ribbons which are limited to four thallium atoms along $\bar{b}$ and are connected to ribbons of four CO_3 groups. Thus, the $[TlO]_x$ layers of the structure $Tl_{0.5}Pb_{0.5}Sr_4Cu_2CO_3O_7$ are replaced by

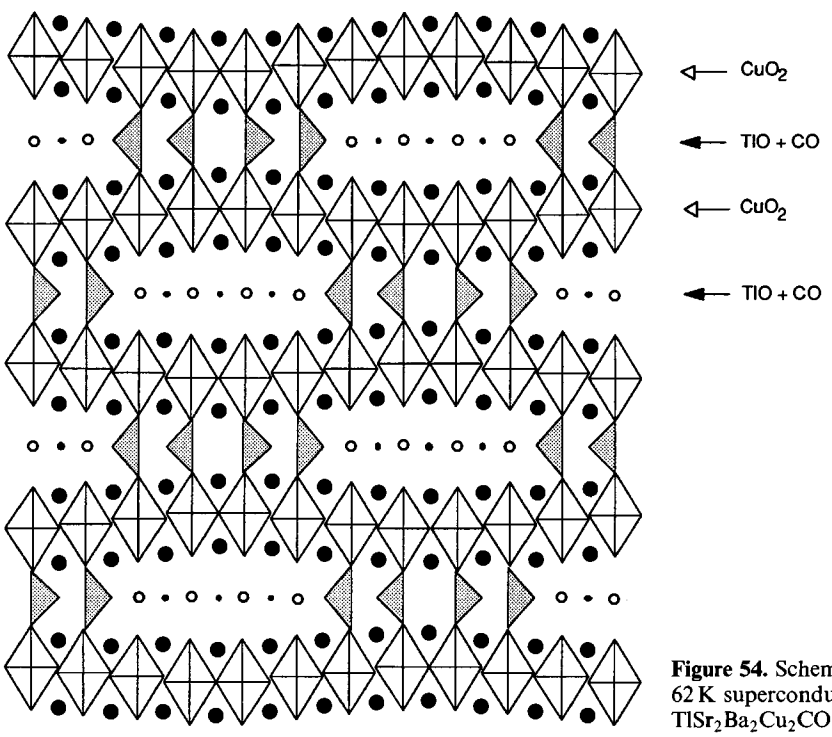


Figure 54. Schematic structure of the 62 K superconductor $TlSr_2Ba_2Cu_2CO_3O_7$.

mixed layers $[(TlO)_4(CO)_4]_x$ characterized by a sequence of four thallium atoms and four carbonate groups along $\bar{b}$. Such a replacement of thallium by carbon implies a strong waving of the octahedral copper layers (Fig. 54) in order to accommodate the differences between the apical Tl–O distances (~2.00 Å) and C–O distances (1.30 Å). Consequently, two successive perovskite layers wave in antiphase so that the larger space between such layers is occupied by thallium, whereas the CO_3 groups are located in the narrower spaces. This waving behavior is clearly observed in the HREM images of Fig. 55; without these convincing observations one would never be able to imagine such a complex and unusual structure. A similar modulation behavior is observed in a number of other related structures such as $HgBa_2Sr_2Cu_2CO_3O_7$ [259] and $Hg_{0.5}Pb_{0.5}Sr_{4-x}Ba_xCu_2(CO_3)O_{7-\delta}$ [260, 261]. In fact, this new structure is also closely related to that of the 1201 thallium cuprates, which consists of single perovskite layers intergrown with rock salt layers. It can be described as a 1201-type structure in which the $[TlO]_x$ layers are replaced by ordered mixed $[(TlO)_4(CO)_4]_4$ layers. One observes again that the critical temperature of this phase is significantly higher than the 1201 mother structures: $TlSr_2CuO_5$ ($T_c = 0$ K) and $TlBa_2CuO_5$ ($T_c = 10$ K), and $TlBaSrCuO_5$ ($T_c = 43$ K). So, it is remarkable that the introduction of CO_3 groups in the 1201 cuprates enhances T_c to 62 K.

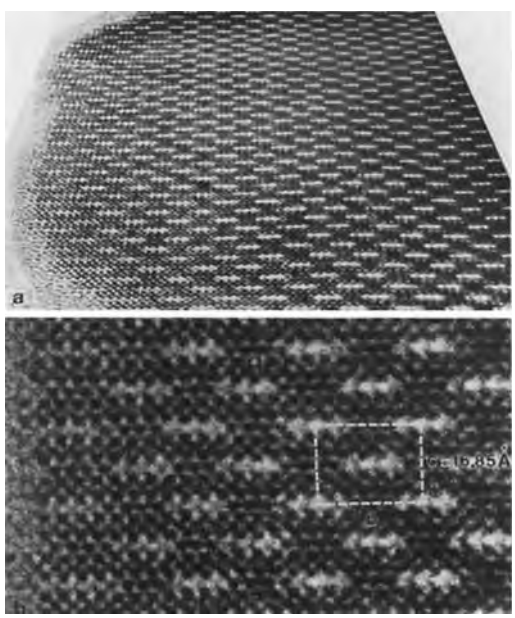


Figure 55. HREM images of modulation in superconducting $Hg_{0.3}Tl_{0.7}Sr_{1.5}Ba_{2.5}Cu_2CO_3O_{7-\delta}$. (a) At lower magnification, under grazing incidence. (b) At high magnification.

Mercury-Based Cuprates

Owing to its preference for twofold coordination, mercury will ensure a junction between either octahedral perovskite layers or pyramidal copper layers, forming linear $Hg^{II}O_2$ groups. The latter layers contain a large number of vacancies, so that they can accommodate the oxygen necessary to create the mixed valence Cu(II)–Cu(III). The great flexibility of the structure of layered cuprates makes it possible to introduce other cations, besides mercury, in the intermediate layers and simultaneously allows incorporation of additional oxygen. After the discovery of the 94 K superconductor $HgBa_2CuO_{4-\delta}$ [262], five series of mercury-based superconducting cuprates have been isolated; four of them belong to the 1201, 1212, 1223, and 1234 structures, respectively, whereas the fifth one is an intergrowth of the 1201 and 2201 structures.

The '1201' Mercury-Based Cuprates: The first oxide of this series that was discovered

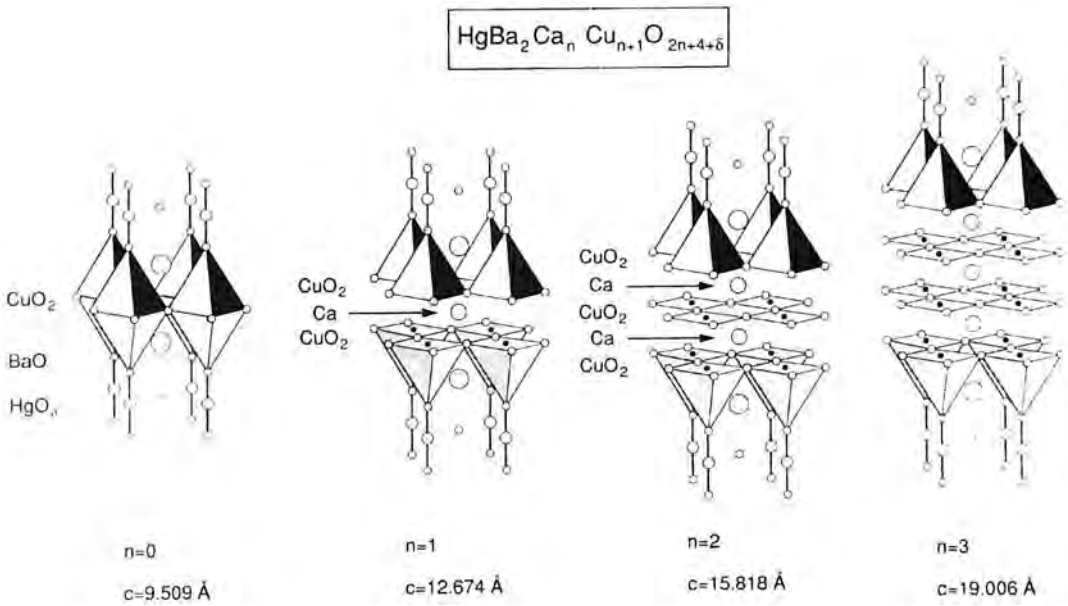


Figure 56. Schematic representation of several members of the structural family $HgBa_2Ca_{n-1}Cu_nO_{2n+2+\delta}$. (a) $n = 1$ or '1201' $HgBa_2CuO_{4+\delta}$. (b) $n = 2$ or '1223' $HgBa_2CaCu_2O_{6+\delta}$. (c) $n = 3$ or '1223' $HgBa_2Ca_2Cu_3O_{8+\delta}$. (d) $n = 4$ or '1223' $HgBa_2Ca_3Cu_4O_{10+\delta}$.

by Putilin et al. [261] can be described as single perovskite octahedral copper layers connected through linear HgO_2 groups (Fig. 56), leading to the ideal formula $HgBa_2CuO_4$. Additional oxygen is then intercalated at the level of the mercury atoms, according to the formula $HgBa_2CuO_{4-\delta}$, which allows an adequate number of holes to be introduced to satisfy the mixed valence Cu(II)–Cu(III). Nevertheless, the mean value of copper of 2.12 ($\delta = 0.06$) is still much smaller than that observed for $YBa_2Cu_3O_7$. This suggests that in this phase the mercury layers have the same role of hole reservoirs as in the bismuth cuprates or the thallium cuprates.

The substitution of bismuth for mercury is possible; it occurs for the particular molar ratio of Hg/Bi = 1. A new superconducting phase with composition $Hg_{0.5}Bi_{0.5}Sr_{2-x}La_xCuO_{5-\delta}$ has indeed been isolated recently [263] for $0 \leqslant x \leqslant 0.75$. As shown from the [100] and [110] HREM images of this phase (Fig. 57), its structure belongs to the 1201 family, that is, it consists of single

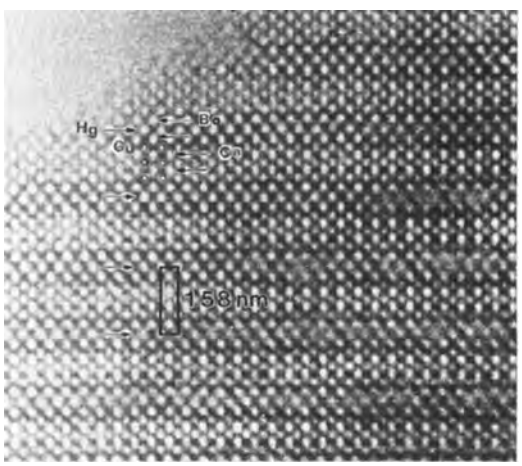


Figure 57. High-resolution image along [100] of $HgBa_2Ca_2Cu_3O_{8+\delta}$. Under this contrast the heavy ions are imaged as bright dots; the imaging code as well as the unit cell are indicated.

perovskite slices $[Sr_{1-x/2}La_{x/2}CuO_3]_x$ intergrown with double rock salt layers $[(Bi_{0.5}Hg_{0.5})(Sr_{1-x/2}La_{x/2})O_{2-\delta}]_x$ (see Fig. 56). Its critical temperature, T_c (onset) $\approx$27 K, is much lower than that of the 1201 barium-based cuprate. Nevertheless, it is interesting to compare its properties to the other 1201 structure cuprates $TlSr_2CuO_{5-\delta}$ and $Tl_{0.5}Pb_{0.5}Sr_2CuO_{5-\delta}$ which do not superconduct. The weaker superconducting properties of this cuprate compared to the pure mercury cuprate $HgBa_2CuO_{4-\delta}$, especially the lower T_c and the broad character of the transition, may be due to the cationic disorder that appears between bismuth and mercury in the rock salt layers. The latter may indeed hinder a perfect coupling between the $[CuO_2]_x$ layers.

The synthesis of the two series of cuprates

$$Hg_{0.3}Pb_{0.7}Sr_{2-\delta}La_xCuO_{4+\delta},\ 0 \leqslant x \leqslant 0.65$$

and

$$Hg_{0.4}Pr_{0.6}Sr_{2-x}Pr_xCu_{4+\delta},\ 0 \leqslant x \leqslant 0.1$$

[261], confirms the ability of strontium to form the 1201 structure. The tetragonal lead cuprate resembles closely the bismuth cuprate, that is, it consists of the stacking of $[(Sr,La)CuO_3]_x$ perovskite layers and $[(Pb,Hg)O_\delta]_x$ rock salt layers with a statistical distribution of lead and mercury. The situation is different for the praseodymium cuprate. The latter exhibits an orthorhombic symmetry, due to a doubling of one '*a*' parameter; this doubling of the unit cell and the ordering between mercury and praseodymium in the $(Hg–Pr)O_\delta$ plane can be detected from HREM images (Fig. 58). The diagnostic volume fraction of the as-synthesized materials is generally weak.

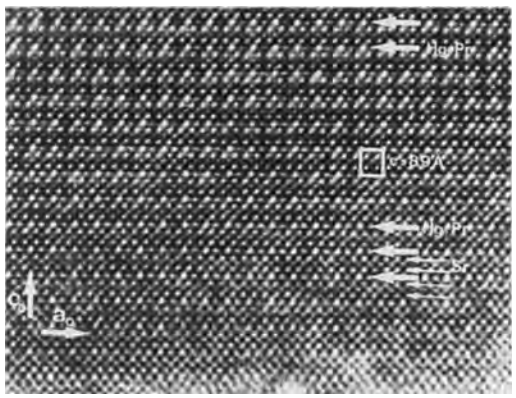


Figure 58. HREM image of $(Hg_{0.4}Pr_{0.6})$-$Sr_{2-x}Pr_xCuO_{4+\delta}$ along [010]. In the thicker areas (upper part of the figure) the Pr–Hg ordering along the *a* axis is clearly observed.

To improve these properties, different treatments are performed. Annealing at 300°C under an oxygen pressure of 20 bar does not modify T_c, nor the diamagnetic volume fraction. The preparation of a deficient lead–mercury phase

$$(Hg_{0.5}Pb_{0.5})_{1-y}Sr_{1.35}La_{0.65}CuO_{4-\delta}$$

leads to the formation of an almost single phase whose parameters are not significantly different from those of the stoichiometric oxide.

The "1212" Mercury-Based Cuprates: After the discovery of superconductivity at 134 K in the system mercury–barium–calcium–copper–oxygen [264], TEM observations allowed the identification of the (poor fraction) of the 134 K superconducting phase, and four mercury-based superconductors with the 1212-type structure have been synthesized almost simultaneously. The first one, $HgBa_2CaCu_2O_{6+\delta}$ [265], exhibits a T_c of 120 K; it can be considered as being derived from the '1-2-3'-type structure. Indeed, the structure of $HgBa_2CaCu_2O_6$

corresponds exactly to the replacement of univalent copper Cu(1) by mercury in the insulating phase $YBa_2Cu_3O_6$. In fact, additional oxygen must be introduced at the level of mercury ($\delta \approx 0.2$) in order to induce superconductivity; the oxygen atoms are distributed over the anionic sites of the rock salt layers in the 1212 structure, so that $HgBa_2CaCu_2O_{6+\delta}$ can also be described as an intergrowth of double pyramidal copper layers and oxygen-deficient rock salt layers. Unfortunately the present structure could only be synthesized under a high pressure of about 50 kbar, so that other routes or other mercury-based cuprates were investigated to reach this critical temperature.

Starting from the idea that the rock salt layer of the 1212 structure may be oxygen-deficient if one introduces mercury, two superconducting 1212-type cuprates were synthesized. The first one, $Y_{0.6}Ca_{0.4}Ba_2Cu_2[Hg_{1-x}M_x]O_{6+\delta}$, with $M = Ca + Cu$ [266], is single phase for $0.4 \leqslant x \leqslant 0.50$, whereas the second one, $Nd_{1-x}Ca_xHg_{1-x'}Cu_{2-x'}O_{6+\delta}$ [267], is almost single phase for $0.4 \leqslant x' \leqslant 0.60$ with $x' = 0$ and 0.2. The X-ray diffraction and HREM study of these oxides show that they both belong to the 1212 series, in agreement with their tetragonal symmetry ($a \approx 3.8$ Å and $c \approx 12.5$ Å). An HREM image along [100] is shown in Fig. 59 for

$$Y_{0.6}Ca_{0.4}Ba_2Cu_2[Hg_{0.6}Cu_{0.2}Ca_{0.2}]O_{6+\delta}$$

These images were extremely useful for the location of the heavy ions. From symmetry considerations as well as from image simulations, one can easily identify the double BaO layers sandwiching the HgO layer. It is clear that the distance between the two BaO layers is larger than the distance between the CuO_2 layers. Most

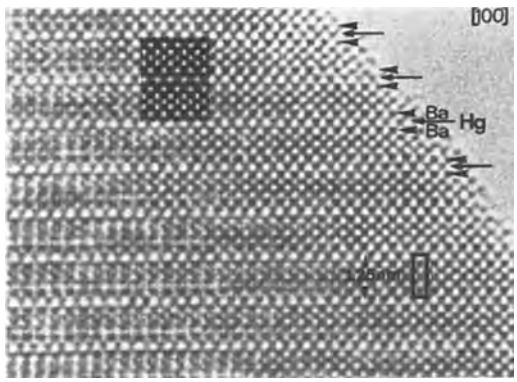


Figure 59. HREM image of $(Y_{0.6}Ca_{0.6})Ba_2Cu_2O_{6+\delta}$ along [100]; the heavy ion configuration is clearly imaged as dark dots in the thinner parts. The 0.387×1.25 nm projected unit cell is indicated. The computer simulated image for a thickness of 3 nm and a defocus value of 60 nm is shown as an inset.

grains are very perfect and almost no (001) planar defects are present. The structure of this phase belongs to the 1212 family, that is, it consists of double pyramidal copper layers intergrown with distorted rock salt layers containing several kinds of cations (mercury, calcium, and copper), according to the general formula

$$[Y_{0.6}Ca_{0.4}]_{FC}Ba_2[Hg_{1-x}M_x]_{RS}Cu_2O_{6+\delta}$$

with $\delta \approx 0.6$ and M = Ca, Cu. The HREM image contrasts observed for the neodymium phase are very similar to those obtained for the yttrium phase, showing that the intermediate rock salt layers are built up from $[HgO_\delta]_x$ or $[Hg_{1-x}Cu_xO_\delta]_x$ slices. The susceptibility curves show that both series of cuprates are superconductors with T_c (onset) values of 90 and 110 K for $Y_{0.6}Ca_{0.4}Ba_2Cu_{2.25}Hg_{0.5}Ca_{0.25}O_{6+\delta}$ and $Nd_{0.6}Ca_{0.4}Ba_2Cu_2HgO_{6+\delta}$, respectively. For both cuprates, the diamagnetic volume fraction of about 25% is characteristic of bulk superconductivity, but in all cases the transition is very broad. This is easily explained by an inhomogeneous

distribution of the cationic species in the rock salt layers; the latter point has been illustrated by EDS analysis, which shows dramatic variations of the cationic distribution from one crystal to another in the 'Nd–Hg'-based cuprates.

The 1212 structure of the mercury-based cuprates can also be stabilized by the introduction, besides mercury, of lead or bismuth. The presence of these cations allows the strontium phase to be synthesized. This is the case for the cuprates $(Pb,Hg)Sr_2(Ca,Y)Cu_2O_{6+\delta}$ [268] and $Hg_{0.5}Bi_{0.5}Sr_2Ca_{1-x}R_xCu_2O_{6+\delta}$ with R = Nd, Y, Pr [269] which exhibit superconductivity up to 90 and 94 K, respectively. The structure of these phases is very similar to that of the barium 1212 cuprates, strontium replacing barium and $[Pb,HgO_\delta]_x$ or $[Bi,HgO_\delta]_x$ layers replacing the $[HgO_\delta]_x$ or $[Hg,CuO_\delta]_x$ layers. They are all characterized by a disordered distribution of mercury and lead or bismuth in the rock salt-type layers. This disordering may be the origin of the broad transition observed for these compounds.

The behavior of the 1212 phase $Bi_{0.5}Hg_{0.5}Sr_2Ca_{1-x}Y_xCu_2O_{6+\delta}$ is very different from that of the 1201 oxide $Bi_{0.5}Hg_{0.5}Sr_2CuO_{4+\delta}$ [263], which does not superconduct. Such a difference between the 1212 and 1201 oxides was previously observed between

$Tl_{0.5}Pb_{0.5}Sr_2(Ca_{1-x}R_x)Cu_2O_7$

and

$Tl_{0.5}Pb_{0.5}Sr_2CuO_5$

[270]. In the same way, the insensitivity to different annealing treatments seems to be a common character of several mercury-based oxides [263, 266, 267, 271].

It is worth pointing out that all the mercury-substituted 1212 cuprates exhibit a smaller critical temperature than the pure mercury cuprate $HgBa_2CaCu_2O_{6+\delta}$. This may be due to the cationic disordering introduced at the level of the mercury layer. Nevertheless, one important feature deals with the fact that all these cuprates, in contrast to $HgBa_2CaCu_2O_{6+\delta}$, can be synthesized at normal pressures

The '2201–1201' Intergrowth $HgTl_2Ba_4Cu_2O_{10+\delta}$: The great similarity between Tl(III) and Hg(II), which are both characterized by a $5d^{10}$ external electronic configuration, suggests that mixed 'Tl–Hg' superconducting cuprates should exist. The superconducting cuprate $HgTl_2Ba_4Cu_2O_{10+\delta}$, [272, 273] confirms this idea. This new phase, which exhibits a tetragonal symmetry ($a = 3.85$ Å and $c = 42.2$ Å), has been detected and studied by HREM. From the [100] image (Fig. 60) one can easily identify the double rows of bright dots with the double thallium

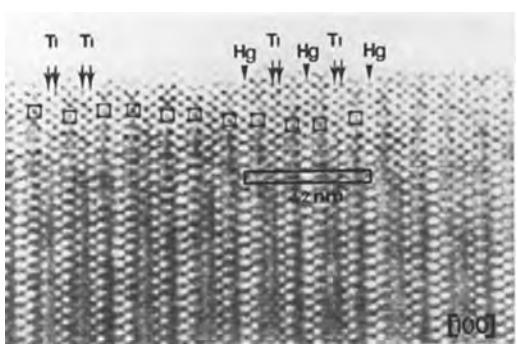


Figure 60. [100] HREM image showing the 4.2 nm unit cell as well as the configuration of the heavy ions. The barium configuration has been indicated by squares, the mercury ions by arrowheads, and the thallium double layers by double arrows. In the left part we observe an occasional defect where the Tl_2–Hg–Tl_2–Hg chain is interrupted by a double Tl_2–Tl_2 sequence.

layers, as previously observed in the 2201 cuprate $Tl_2Ba_2CuO_6$ [272], whereas the single row of bright dots can be identified with the thallium or mercury monolayers observed for the 1201 cuprates, $TlSr_2CUO_5$, or $HgBa_2CuO_4$. They are indicated in Fig. 60 with double arrows and single arrowheads, respectively. The square barium configuration, imaged as dark dots in the thinnest part of the sample, is identified in Fig. 60. From this study, it clearly appears that the ideal structure of the new phase $HgTl_2Ba_4Cu_2O_{10+\delta}$, represented in Fig. 61, consists of an intergrowth of the 2201

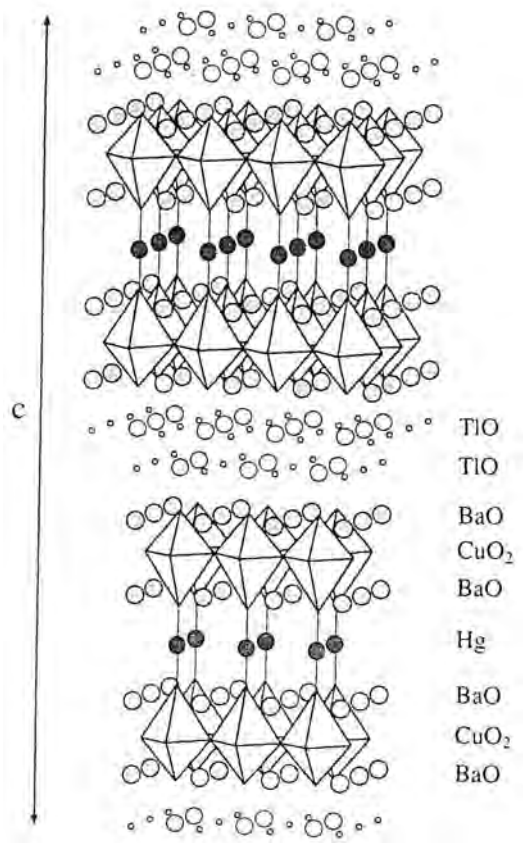


Figure 61. Schematic representation of the idealized structure of $Tl_2HgBa_4Cu_2O_{10}$.

and the 1201 structures, that is, it is built up from single octahedral perovskite layers intergrown with distorted triple and double rock salt-type layers alternately. Taking into consideration the observed molar ratio Tl/Hg = 2, the higher stability of the 2201 phase $Tl_2Ba_2CuO_6$ with respect to the 1201 phase $TlBa_2CuO_5$, and the ability of mercury to form the 1201 phase, it appears most likely that the 2201 slices are mainly built up from thallium bilayers, whereas the 1201 slices contain mainly mercury. Nevertheless, partial mixed occupation of these two kinds of cation sites cannot be ruled out. As a consequence, the structure of this phase can be represented by the formula

$$[Tl_{2-\delta}Hg_\delta Ba_2CuO_6]^{2201}[Hg_{1-\delta}Tl_\delta Ba_2CuO_{4+\delta}]^{1201}$$

The '1223', '1234' and '12nn+1' Cuprates: Recently, several higher-order members of the mercury-based superconducting homologous family $HgBa_2Ca_{n-1}Cu_nO_{2n+2+\delta}$ have been synthesized in more or less pure form [274]. The critical temperature (onset of the Meissner signal) of these materials is known to increase from 94 K (for the $n = 1$ compound) to 126 K for the $n = 2$ compound, and even to a maximum of $T_c = 133.5$ K for the $n = 3$ member. For larger n values, T_c seems to decrease [275], similar to the behavior of the related thallium compounds [276].

The structure of all members is similar; they contain rock salt-like slabs [(BaO)(HgO_δ)(BaO)] alternating with perovskite slabs of the type $[(CuO_2)(Ca)]_{n-1}(CuO_2)$. The existence and the structure of the $n = 1$ member of this family was first reported by Putilin

et al. [262]. Samples containing sequences of the higher-order members were later found by Schilling et al. [264]. HREM not only allows study of the perfect structure of higher-order members of the mercury family but also investigation of possible defects in such compounds. High-quality high-resolution images for the lower members of the $12(n-1)n$ family have been published [277, 278].

Electron diffraction images along several reciprocal zone axes show that some materials include crystallites corresponding to n values ranging from 2 to 7. The higher members of this series cannot be obtained in a pure form so that X-ray studies are extremely difficult, if not impossible. A compilation of [100] diffraction patterns corresponding to these n values is shown in Fig. 62.

Real space observations allow the investigation of the perfection of the cyrstals. For the $n = 1$, $n = 2$, and $n = 3$ compounds, hardly any deviation from perfection is found, and virtually no secondary phases are met. An example of a [100] high-resolution image for the 1223 compound is shown in Fig. 57; the cation configuration can be readily identified from symmetry considerations. There is a one-to-one correspondence between the white dot configuration of the HREM image and the cation configuration. Such semi-intuitive interpretations have of course to be confirmed by computer simulations based on the structure model as determined by neutron diffraction and knowledge of the electron microscope parameters [278]. Such calculated HREM images for different thicknesses and different defocus values allow us to analyze in detail the atomic structure of planar defects in the present compounds.

The most common defect encountered, particularly for higher-order members of the family, is the intergrowth of different n

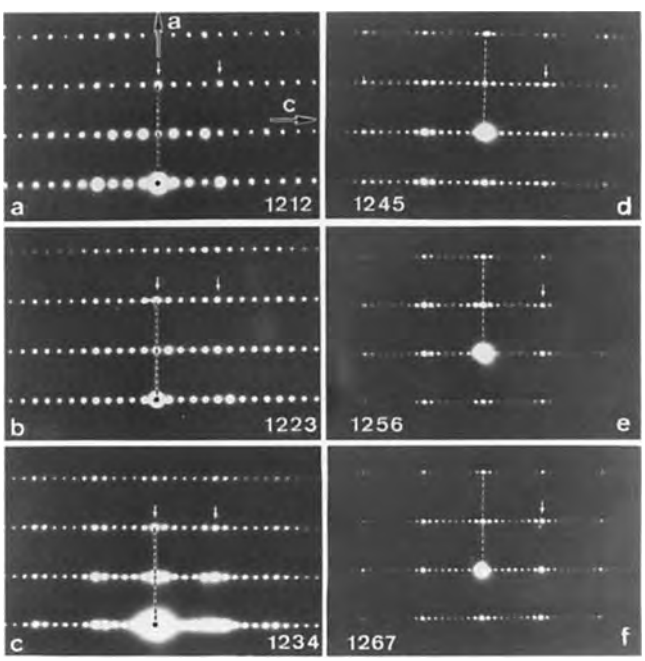


Figure 62. [100] electron diffraction patterns for different members of the series with n ranging from (a) $n = 2$ to (f) $n = 7$.

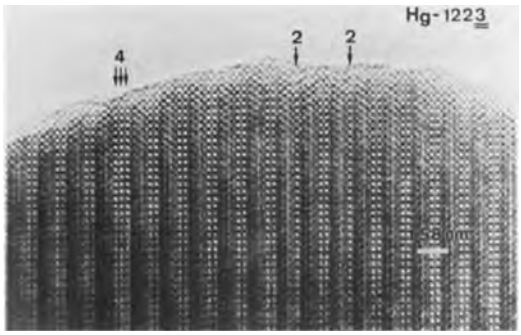


Figure 63. Occasional stacking errors in a mercury 1223 crystal; locally, mercury 1212 (indicated by the numeral 2) and mercury 1234 (indicated by the numeral 4) are found. The calcium configuration shows up here as very pronounced dots, compatible with a focus value of around −200 Å.

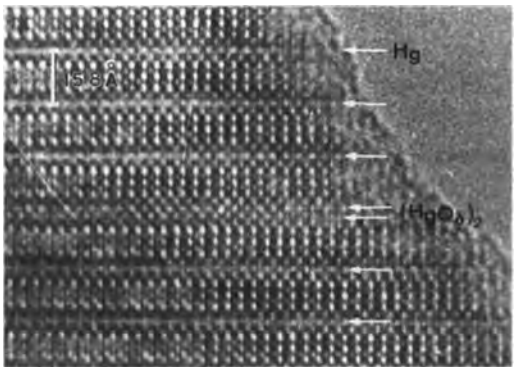


Figure 64. HREM image of an isolated defect in a mercury 1223 crystal, where a double layer of HgO_δ is present, creating locally the 2223 structure.

members. Generally, such defects are isolated, occurring in an otherwise perfect crystal. Examples are shown in Fig. 63 where in a 1223 crystal locally two slabs of 1212 (indicated by the numeral 2, referring to the number of copper layers in the structure) and one slab of 1234 (indicated by the numeral 4) are found. With the help of the image simulations, which allow us to identify the extra bright dot configuration in Fig. 63 as the calcium layers, such defects can readily be analyzed. For materials corresponding to n values exceeding 4, these defects are a common feature.

A most remarkable defect, however, is shown in Fig. 64. The image contrast is such that the copper ions are imaged as bright dots, while the mercury plane is imaged as a more or less diffuse line with maxima at the projected atom positions. It is evident from the imaging code that locally we have four rock salt layers rather than the usual sequence of three layers [(BaO)(HgO_δ)(BaO)]. A single mercury layer is replaced by a double mercury layer of bright dots. Simulated images of a trial structure with a double (HgO_δ) layer in the structure confirm the present interpretation.

The occurrence of such defects with a double (HgO_δ) layer suggests the feasibility of producing '$22n-1n$'-type materials. Compounds such as Hg–22 $n-1$ n can only be stabilized if the oxygen sites of the mercury layers are occupied by more than 50%. The neutron diffraction refinements of the different mercury members show that the largest occupation factors for those oxygen atoms is at most 40%. However, by replacing some of the mercury by other elements, such as praseodymium or copper, the '2212' structure with a double (HgO_δ) layer can be stabilized [279].

In some of the preparations (for details, see Kopnin et al. [280]), the critical temperature T_c was much lower than the above-reported values. With the help of microanalysis and HREM one can find that some of the mercury positions are occupied by carbonate groups but in a very disordered way. Electron diffraction, which is very sensitive to local ordering or short-range ordering, does not show any

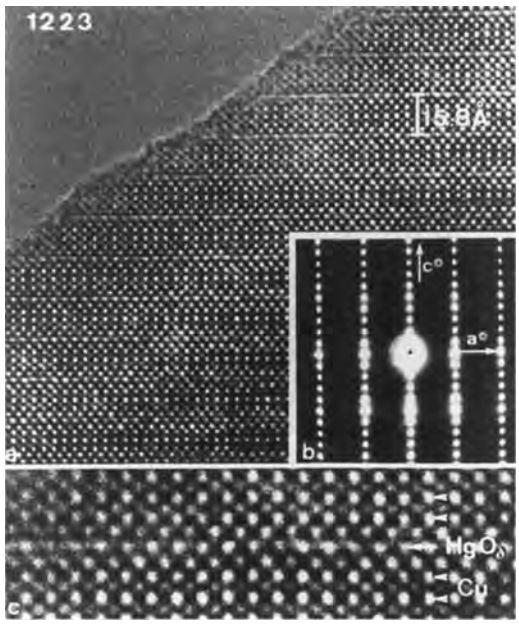


Figure 65. (a) HREM image along [010] of a mercury 1223 compound, showing the individual cation columns as bright dots. (b) The corresponding electron diffraction pattern. (c) Higher magnification of part (a); note that in the HgO_δ layer the intensity of the individual dots is not homogeneous.

diffuse scattering (Fig. 65b). Also, the corresponding HREM image of Fig. 65a does not show any superstructure formation. The HgO_δ layer can be easily identified as the more diffuse layer; along this layer, however, we do see variations in the intensities of the different bright dots (see Fig. 65c, which is an enlargement of Fig. 65a). One can quantify these variations by making densitometer traces along the HgO_δ layer as well as along one of the neighboring CuO_2 layers. The results are shown in Fig. 66, where we have plotted the intensity for different bright dots along a single HgO_δ row; there is clearly a pseudo-periodic variation in the intensity; such variation is not observed in the corresponding measurements for the BaO or the CuO_2 layer (Fig. 66b). These intensity variations are correlated with the occupation of the mercury sublattice, which from detailed EDX measurements is only $Hg_{0.7-0.8}$. The presence of carbonate groups cannot be verified by EDX, because carbon is at the limit of detection and contamination will always introduce carbon.

Acknowledgements

I am grateful to the many scientists allowing me to use their illustrations. Most of the work within the EMAT group has been done in collaboration with T. Krekels, S. Amelinckx, and H. W. Zandbergen.

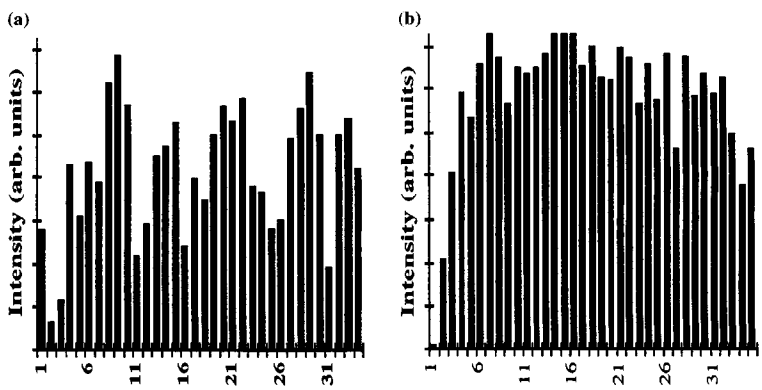


Figure 66. Intensity measurement of individual dots from Fig. 64; (a) along the HgO_δ layer and (b) along the neighbouring CuO_2 layer.

8.4 References

[1] J. G. Bednorz, K. A. Müller, *Z. Physik. B* **1986,** *64*, 189.

[2] A. Schilling, M. Cantoni, J. D. Guo, H. R. Ott, *Nature* **1993,** *363*, 56.

[3] M. K. Wu, J. R. Asburn, C. J. Torng, P. H. Hor, R. L. Meng, L. Gao, Z. J. Huang, C. W. Chu, *Phys. Rev. Lett.* **1987,** *58*, 908.

[4] M. Morita, K. Miyamoto, K. Doi, M. Murakami, K. Sawano, S. Matsuda, *Physica C* **1990,** *172*, 383.

[5] Y. L. Chen, H. M. Chan, M. P. Harmer, V. R. Todt, S. Sengupta, D. Shi, *Physica C* **1994,** *234*, 232.

[6] F. Frangi, E. Varesi, G. Ripamonti, S. Zannella, *Physica C* **1994,** *233*, 301.

[7] M. A. Rodriguez, B.-J. Chen, R. L. Snyder, *Physica C* **1992,** *195*, 185.

[8] A. Goyal, P. D. Funkenbusch, D. M. Kroeger, S. J. Burns, *Physica C* **1991,** *182*, 203.

[9] B. J. Chen, M. A. Rodriguez, S. T. Misture, R. L. Snyder, *Physica C* **1994,** *217*, 367.

[10] A. Otto, R. Kontra, J. B. Vander Sande, *Physica C* **1992,** *190*, 581.

[11] I. Monot, T. Higuchi, N. Sakai, M. Murakami, *Physica C* **1994,** *233*, 155.

[12] H. Komatsu, Y. Kato, S. Miyashita, T. Inoue, S. Hayashi, *Physica C* **1991,** *190*, 14.

[13] S. R. Su, M. O'Connor, M. Levinson, P. G. Rossoni, *Physica C* **1991,** *178*, 81.

[14] C. T. Lin, S. X. Li, A. Mackenzie, W. Zhou, P. D. Hunneyball, W. Y. Liang, *Physica C* **1992,** *193*, 129.

[15] C. T. Lin, W. Y. Liang, *Physica C* **1994,** *225*, 275.

[16] P. E. D. Morgan, T. J. Doi, R. M. Housley, *Physica C* **1993,** *213*, 438.

[17] Y. F. Li, Z. Z. Sheng, S. Sengupta, K. C. Goretta, P. E. D. Morgan, *Physica C* **1994,** *232*, 184.

[18] W. Zhang, E. E. Hellstrom, *Physica C* **1993,** *218*, 141.

[19] L. E. C. Van de Leemput, P. J. M. Van Bentum, L. W. M. Schreurs, H. Van Kempen, *Physica C* **1988,** *152*, 99.

[20] S. Tanaka, E. Ueda, M. Sato, *Physica C* **1994,** *224*, 126.

[21] T. Kroener, G. Linker, O. Meyer, B. Strehlau, Th. Wolf, H. J. Hug, T. A. Jung, H.-J. Güntherodt, *Physica C* **1992,** *191*, 243.

[22] M. Hawley, I. D. Raistrick, J. G. Beery, R. J. Houlton, *Science* **1991,** *251*, 1587.

[23] Ch. Gerber, D. Anselmetti, J. G. Bednorz, J. Mannhart, D. G. Schlom, *Nature* **1991,** *350*, 279.

[24] L. Luo, M. E. Hawley, C. J. Maggiore, R. C. Dye, R. E. Muenchhausen, *Appl. Phys. Lett.* **1993,** *62*, 485.

[25] H. U. Krebs, Ch. Kraus, *Appl. Phys. Lett.* **1991,** *59*, 2180.

[26] I. Heyvaert, E. Osqueguil, C. V. Van Haesendonck, Y. Bruynseraede, *Appl. Phys. Lett.* **1992,** *61*, 111.

[27] Xing Zhu, G. C. Xiong, R. Lui, Y. J. Li, G. J. Lian, J. Li, Z. Z. Gan, *Physica C* **1993,** *216*, 153.

[28] M. Boekholt, M. Fleuster, F. Nouvertné, M. Herrmann, G. Güntherodt, C. Jaekel, D. Anselmetti, H. J. Güntherodt, D. Wehler, G. Müller, *Physica C* **1992,** *203*, 180.

[29] A. Inoue, H. Mukaida, M. Nakao, R. Yoshizaki, *Physica C* **1994,** *233*, 49.

[30] M. D. Kirk, J. Nogami, A. A. Baski, D. B. Mitzi, A. Kapitulnik, T. H. Geballe, C. F. Quate, *Science* **1988,** *242*, 1673.

[31] C. K. Shih, R. M. Feenstra, G. V. Chandrashekar, *Phys. Rev. B* **1991,** *43*, 7913.

[32] V. P. S. Awana, S. K. Agarwal, R. Ray, S. Gupta, A. V. Narlikar, *Physica C* **1992,** *191*, 43.

[33] G. Van Tendeloo, H. W. Zandbergen, S. Amelinckx, *Solid State Commun.* **1987,** *63*, 389.

[34] J. C. Barry, *J. Electron Microsc. Techniques* **1988,** *8*, 325.

[35] S. Iijima, T. Ichihashi, Y. Kubo, J. Tabuchi, *Jpn. J. Appl. Phys.* **1987,** *26*, L1790.

[36] J. L. Hodeau, P. Bordet, J.-J. Capponi, C. Chaillout, M. Marezio, *Physica C* **1988,** *153–155*, 582.

[37] E. A. Hewat, M. Dupuy, A. Bourret, J.-J. Capponi, M. Marezio, *Solid State Commun.* **1987,** *64*, 517.

[38] C. J. Jou, J. Washburn, *J. Mater. Res.* **1989,** *4*, 795.

[39] G. Roth, G. Heger, B. Renker, J. Pannetier, V. Caignaert, M. Hervieu, B. Raveau, *Z. Phys. B* **1988,** *71*, 43.

[40] B. D. Dunlop, J. D. Jorgensen, C. Segre, A. E. Dwight, J. L. Matykiewicz, H. Lee, W. Peng, C. W. Kimball, *Physica C* **1989,** *158*, 397.

[41] Y. Xu, M. Suenaga, J. Tafto, R. L. Sabatini, A. R. Moodenbaugh, P. Zolliker, *Phys. Rev. B* **1989,** *39*, 6667.

[42] E. Suard, V. Caignaert, A. Maignan, F. Bourée, B. Raveau, *Physica C* **1993,** *205*, 63.

[43] E. Suard, A. Maignan, V. Caignaert, B. Raveau, *Physica C* **1992,** *200*, 43.

[44] T. Krekels, G. Van Tendeloo, D. Broddin, S. Amelinckx, L. Tanner, M. Mehbod, E. Vanlathem, R. Deltour, *Physica C* **1991,** *173*, 361.

[45] G. Van Tendeloo, S. Amelinckx, *J. Electron Microsc. Techniques* **1988,** *8*, 285.

[46] H. W. Zandbergen, G. Van Tendeloo, T. Okabe, S. Amelinckx, *Phys. Status Solidus A* **1987,** *103*, 45.
[47] G. Van Tendeloo, D. Broddin, H. W. Zandbergen, S. Amelinckx, *Physica C* **1990,** *167*, 627.
[48] S. Amelinckx, *Dislocations in Solids* **1979,** part II, (Ed.: F. R. N. Nabarro), North-Holland, Amsterdam p. 327.
[49] U. Welp, M. Grimsditch, H. You, W. H. Kwok, M. M. Fang, G. W. Crabtree, J. Z. Liu, *Physica C* **1989,** *161*, 1.
[50] H. W. Zandbergen, R. Gronsky, K. Wang, G. Thomas, *Nature* **1988,** *331*, 596.
[51] D. J. Li, H. Shibahara, J. P. Zhang, L. D. Marks, H. O. Marcy, S. Song, *Physica C* **1988,** *156*, 201.
[52] B. Domenges, M. Hervieu, C. Michel, B. Raveau, *Europhys. Lett.* **1987,** *4*, 211.
[53] D. E. Morris, N. G. Asmar, J. H. Nickel, R. L. Sid, J. Y. T. Wei, J. E. Post, *Physica C* **1989,** *159*, 287.
[54] J. Karpinski, E. Kaldis, E. Jilek, S. Rusiecki, B. Bucher, *Nature* **1988,** *336*, 660.
[55] P. Bordet, C. Chaillout, J. Chenavas, J. L. Hodeau, M. Marezio, J. Karpinski, E. Kaldis, *Nature* **1989,** *334*, 596.
[56] T. Krekels, G. Van Tendeloo, S. Amelinckx, J. Karpinski, S. Rusiecki, E. Kaldis, E. Jilek, *Physica C* **1991,** *178*, 383.
[57] T. Krekels, G. Van Tendeloo, S. Amelinckx, J. Karpinski, E. Kaldis, S. Rusiecki, *Solid State Commun.* **1991,** *79*, 607.
[58] R. Ramesh, S. Jin, P. March, *Nature* **1990,** *346*, 420.
[59] R. Beyers, B. T. Ahn, G. Gorman, V. Y. Lee, S. S. P. Parkin, M. L. Ramirez, K. P. Roche, J. E. Vasquez, T. M. Gür, R. A. Huggings, *Nature* **1989,** *340*, 619.
[60] H. F. Poulsen, N. H. Andersen, J. V. Andersen, H. Bohr, O. G. Mauritsen, *Nature* **1991,** *349*, 594.
[61] T. Krekels, H. Zou, G. Van Tendeloo, D. Wagener, M. Buchgeister, S. M. Hosseini, P. Herzog, *Physica C* **1992,** *196*, 363.
[62] T. Krekels, G. Van Tendeloo, D. Broddin, S. Amelinckx, L. Tanner, M. Mehbod, E. Vanlathem, R. Deltour, *Physica C* **1991,** *173*, 361.
[63] W. W. Schmahl, A. Punis, E. Salje, P. Freeman, A. Graeme-Barber, R. Jones, K. K. Singh, J. Blunt, P. P. Edwards, J. Loram, K. Mirza, *Phil. Mag. Lett.* **1989,** *60*, 341.
[64] R. J. Cava, A. W. Hewat, E. A. Hewat, B. Batlogg, M. Marezio, K. M. Rabe, J. J. Krajewski, W. F. Peck Jr., L. W. Rupp Jr., *Phys. C* **1989,** *165*, 419.
[65] A. Ourmazd, J. C. H. Spence, *Nature* **1987,** *329*, 425.
[66] Y. P. Lin, J. E. Greedan, A. H. O'Reilly, J. N. Reimers, C. V. Stager, M. L. Post, *J. Sol. Stat. Chem.* **1990,** *84*, 226.
[67] C. Chaillout, M. A. Alario Franco, J. J. Capponi, J. Chenavas, J. L. Hodeau, M. Marezio, *Phys. Rev. B* **1987,** *36*, 7118.
[68] Y. Zhu, A. R. Moodenbaugh, M. Suenaga, J. Tafto, *Physica C* **1990,** *167*, 363.
[69] C. H. Chen, D. J. Werder, L. F. Schneemeyer, P. K. Gallagher, J. V. Waszczak, *Phys. Rev. B* **1988,** *38*, 2888.
[70] C. N. R. Rao, R. Nagarajan, A. K. Ganguli, G. N. Subbana, S. V. Bhat, *Phys. Rev. B* **1990,** *42*, 6765.
[71] C. P. Burmester, S. Quong, L. T. Wille, R. Gronsky, B. T. Ahn, V. Y. Lee, R. Beyers, T. M. Gür, R. A. Huggins, *Mater. Res. Soc. Symp. Proc.* **1990,** *183*, 369.
[72] J.-L. Hodeau, P. Bordet, J.-J. Capponi, C. Chaillout, M. Marezio, *Physica C* **1988,** *153–155*, 582.
[73] R. Beyers, B. T. Ahn, G. Gorman, V. Y. Lee, S. S. P. Parkin, M. L. Ramirez, K. P. Roche, J. E. Vazquez, T. M. Gür, R. A. Huggings, *Nature* **1989,** *340*, 619.
[74] M. A. Alario-Franco, C. Chaillout, J. J. Capponi, J. Chenavas, *Mater. Res. Bull.* **1987,** *22*, 1685.
[75] M. A. Alario-Franco, C. Chaillout, J. J. Capponi, J. Chenavas, M. Marezio, *Phys. C* **1988,** *156*, 455.
[76] C. J. Hou, A. Manthiram, L. Rabenberg, J. B. Goodenough, *J. Mater. Res.* **1990,** *5*, 9.
[77] C. Chaillout, M. A. Alario Franco, J. J. Capponi, J. Chenavas, P. Strobel, M. Marezio, *Solid State Commun.* **1987,** *65*, 283.
[78] M. Hervieu, B. Domengès, B. Raveau, M. Post, W. R. McKinnon, J. M. Tarascon, *Mater. Lett.* **1989,** *8*, 73.
[79] M. Hervieu, B. Domengès, C. Michel, J. Provost, B. Raveau, *J. Solid State Chem.* **1987,** *71*, 263.
[80] G. Van Tendeloo, H. W. Zandbergen, S. Amelinckx, *Solid State Commun.* **1987,** *63*, 603.
[81] R. M. Fleming, L. F. Schneemeyer, P. K. Galagher, B. Batlogg, L. W. Rupp, J. V. Waszczak, *Phys. Rev. B* **1988,** *37*, 7920.
[82] H. You, J. D. Axe, X. B. Kan, S. Hashimoto, S. C. Moss, J. Z. Liu, G. W. Crabtree, D. J. Lam, *Phys. Rev. B* **1988,** *38*, 9213.
[83] T. Zeiske, D. Hohlwein, R. Sonntag, F. Kubanek, T. Wolf, *Phys. C* **1992,** *194*, 1.
[84] Y. Nakazawa, M. Ishikawa, *Phys. C* **1989,** *158*, 381.
[85] Y. Nakazawa, M. Ishikawa, *Phys. C* **1989,** *162–164*, 83.
[86] Y. Ueda, K. Kosuge, *Phys. C* **1988,** *156*, 281.

[87] R. J. Cava, B. Batlogg, C. H. Chen, E. A. Rietman, S. M. Zahurak, D. Werder, *Phys. Rev. B* **1987,** *36*, 5719.
[88] H. Strauven, J. P. Locquet, O. B. Verbeke, Y. Bruynseraede, *Solid State Commun.* **1987,** *65*, 293.
[89] P. Gerdanian, C. Picard, B. Touzelin, *Physica C* **1991,** *182*, 11.
[90] D. Van Dyck, C. Condé, S. Amelinckx, *Phys. Status Solid A* **1979,** *56*, 327.
[91] T. Krekels, G. Van Tendeloo, S. Amelinckx, D. Wagener, M. Buchgeister, S. M. Hosseini, P. Herzog, *Physica C* **1992,** *196*, 363.
[92] S. Yang, H. Claus, B. W. Veal, R. Wheeler, A. P. Paulikas, J. W. Downey, *Physica C* **1992,** *193*, 243.
[93] S. Rusiecki, B. Bucher, E. Kaldis, E. Jilek, J. Karpinski, C. Rossel, B. Pümpin, H. Keller, W. Kündig, T. Krekels, G. Van Tendeloo, *J. Less Common Metals* **1990,** *164–165*, 31.
[94] J. Reyes-Gasga, T. Krekels, G. Van Tendeloo, J. Van Landuyt, S. Amelinckx, W. H. M. Brugginck, H. Verweij, *Physica C* **1989,** *159*, 831.
[95] G. Van Tendeloo, S. Amelinckx, *J. El. Mic. Technique* **1988,** *8*, 285.
[96] A. G. Khachaturyan, J. W. Morris, *Phys. Rev. Lett.* **1988,** *61*, 215.
[97] D. de Fontaine, G. Ceder, M. Asta, *Nature* **1990,** *343*, 544.
[98] H. W. Zandbergen, Y. K. Huang, M. J. V. Menken, J. N. Li, K. Kadowaki, A. A. Menovsky, G. Van Tendeloo, S. Amelinckx, *Nature* **1988,** *332*, 620.
[99] H. Maeda, Y. Tanaka, M. Eukutomi, T. Asano, *Jpn. J. Appl. Phys.* **1987,** *27*, L209.
[100] C. Michel, M. Hervieu, M. M. Borel, A. Grandin, F. Deslandes, J. Provost, B. Raveau, *Z. Physik B* **1987,** *68*, 421.
[101] Z. Z. Sheng, A. M. Hermann, *Nature* **1988,** *332*, 138.
[102] Y. Hirotsu, O. Tomioka, T. Ohkubo, N. Yamamoto, Y. Nakamura, S. Nagakura, T. Komatsu, K. Matsushita, *Jpn. J. Appl. Phys.* **1988,** *27*, 1869.
[103] Y. Hirotsu, N. Yamamoto, O. Tomioka, K. Miyagawa, Y. Nakamura, Y. Inoue, S. Nagakura, Y. Iwai, M. Takata in *Advances in Superconductivity II* (Eds.: T. Ishiguro, K. Kajimura), Springer-Verlag, Tokyo **1990,** p. 161.
[104] Z. Hiroi, Y. Ikeda, M. Takano, Y. Bando, *J. Mater. Res.* **1991,** *6*, 435.
[105] T. Onozuka, T. Kajitani, M. Hirabayashi, H. Sato, T. E. Mitchell, *Jpn. Appl. Phys.* **1989,** *28*, 1775.
[106] O. Eibl, *Physica C* **1990,** *168*, 249.
[107] O. Eibl, *Physica C* **1990,** *169*, 441.
[108] H. W. Zandbergen, W. A. Groen, G. Van Tendeloo, J. Van Landuyt, S. Amelinckx, *Solid State Commun.* **1988,** *66*, 397.
[109] G. Van Tendeloo, H. W. Zandbergen, S. Amelinckx, *Solid State Commun.* **1988,** *66*, 927.
[110] G. Van Tendeloo, H. W. Zandbergen, J. Van Landuyt, S. Amelinckx, *Appl. Phys. A* **1988,** *46*, 153.
[111] O. Eibl, *Physica C* **1990,** *168*, 215.
[112] O. Eibl, *Physica C* **1991,** *175*, 419.
[113] Y. Hirotsu, O. Tomioka, N. Yamamoto, Y. Nakamura, S. Nagakura, Y. Iwai, M. Takata, *Jpn. Appl. Phys.* **1989,** *28*, 1783.
[114] H. W. Zandbergen, W. A. Groen, G. Van Tendeloo, S. Amelinckx, *Appl. Phys. A* **1989,** *48*, 305.
[115] Y. Inoue, Y. Shichi, F. Munakata, M. Yamanaka, *Phys. Rev. B* **1989,** *40*, 7307.
[116] T. Onozuka, Y. Hirotsu, *Acta Crystallogr. Ser. A* **1994,** *50*, 231.
[117] K. Yanagisawa, Y. Matsui, T. Hasegawa, T. Koizumi, S. Takekawa, *Physica C*, *222*, **1994,** 184–190.
[118] K. Yanagisawa, Y. Matsui, T. Hasegawa, T. Koizumi, *Physica C* **1993,** *208*, 51.
[119] K. Yanagisawa, Y. Matsui, K. Shoda, E. Takayama-Muromachi, S. Horiuchi, *Physica C* **1992,** *196*, 34.
[120] H. W. Zandbergen, W. A. Groen, F. C. Mijlhoff, G. Van Tendeloo, S. Amelinckx, *Physica C* **1988,** *156*, 325.
[121] H. Heinrich, G. Kostorz, B. Heeb, L. J. Gauckler, *Physica C* **1994,** *224*, 133.
[122] A. Yamamoto, M. Onoda, E. Takayama-Muromachi, F. Izumi, *Phys. Rev. B* **1990,** *42*, 4228.
[123] A. I. Beskrovnyi, M. Dlouha, Z. Jirak, S. Vratislav, E. Pollert, *Physica C* **1990,** *166*, 79.
[124] A. I. Beskrovnyi, M. Dlouha, Z. Jirak, S. Vratislav, *Physica C* **1990,** *171*, 19.
[125] J. M. Tarascon, Y. LePage, W. R. McKinnon, R. Ramesh, M. Eibschutz, E. Tselepis, E. Wang, G. W. Hull, *Physica C* **1990,** *167*, 20.
[126] Y. Le Page and W. R. McKinnon, J.-M. Tarascon, P. Barboux, *Phys. Rev. B* **1989,** *40*, 6810.
[127] G. Calestani, M. G. Francesconi, G. Salsi, G. D. Andretti, A. Migliori, *Physica C* **1992,** *197*, 283.
[128] G. Calestani, G. Salsi, M. G. Francesconi, M. Masini, L. Dimesso, A. Migliori, X. F. Zhang, G. Van Tendeloo, *Physica C* **1993,** *206*, 33.
[129] A. Q. Pham, A, Maignan, M. Hervieu, C. Michel, J. Provost, B. Raveau, *Physica C* **1992,** *191*, 77.

[130] S. S. P. Parkin, V. Y. Lee, A. I. Nazzal, R. Savoy, R. Beyers, *Phys. Rev. Lett.* **1988,** *61*, 750.
[131] M. Verwerft, G. Van Tendeloo, S. Amelinckx, *Physica C* **1988,** *156*, 607.
[132] S. Iijima, T. Ishihashi, Y. Shimakawa, T. Manako, *Jpn. Appl. Phys.* **1988,** *27*, 1054.
[133] G. Van Tendeloo, S. Amelinckx, *Physica C* **1991,** *176*, 575.
[134] G. Van Tendeloo, S. Amelinckx, *Appl. Phys. A* **1991,** *52*, 313.
[135] M. Verwerft, G. Van Tendeloo, D. G. Hinks, B. Dabrowski, D. R. Richards, A. W. Mitchell, D. T. Marx, Shiyou Pei, J. D. Jorgensen, *Phys. Rev. B* **1991,** *44*, 9547.
[136] C. Chaillout, J. P. Remeika, *Solid State Commun.* **1985,** *56*, 833.
[137] S. Pei, N. J. Zaluzec, J. D. Jorgensen, B. Dabrowski, D. G. Hinks, A. W. Mitchell, D. R. Richards, *Phys. Rev. B* **1989,** *39*, 811.
[138] S. Pei, J. D. Jorgensen, B. Dabrowski, D. G. Hinks, D. R. Richards, A. W. Mitchell, J. M. Newsam, S. K. Sinha, D. Vaknin, A. J. Jacobson, *Phys. Rev. B* **1990,** *41*, 4126.
[139] E. A. Hewat, C. Chaillout, M. Godinho, M. F. Gorius, M. Marezio, *Physica C* **1989,** *157*, 228.
[140] R. P. Huebener, *Magnetic Flux Structures in Superconductors* **1978,** Springer-Verlag, New York.
[141] A. Tonomura, *Physica C* **1994,** *235–240*, 33.
[142] H. F. Hess, R. B. Robinson, R. C. Dynes, J. M. Valles, J. V. Wasczcak, *Phys. Rev. Lett.* **1989,** *62*, 214.
[143] A. Moser, H. J. Hug, I. Parashikov, B. Stiefel, O. Fritz, H. Thomas, A. Baratoff, H.-J. Güntherodt, P. Chaudhari, *Phys. Rev. Lett.* **1995,** *74*, 1847.
[144] H. J. Hug, A. Moser, I. Parashikov, B. Stiefel, O. Fritz, H.-J. Güntherodt, H. Thomas, *Physica C* **1994,** *235–240*, 2695.
[145] Th. Schuster, M. Leghissa, M. R. Koblischka, B. H. Kuhn, M. Kraus, H. Kronmüller, G. Szemann-Ischenko, *Physica C* **1992,** *203*, 203.
[146] M. V. Indenbom, A. Forkl, B. Ludescher, H. Kronmüller, H.-U. Habermeier, B. Leibold, G. D'Anna, T. W. Li, P. H. Kes, A. A. Menovsky, *Physica C* **1994,** *226*, 325.
[147] M. R. Koblischka, Th. Schuster, H. Kronmüller, *Physica C* **1994,** *219*, 205.
[148] M. Turchinskaya, D. L. Kaiser, F. W. Gayle, A. J. Shapiro, A. Roytburd, L. A. Dorosinskii, V. I. Nikitenko, A. A. Polyanskii, V. K. Vlasko-Vlasov, *Physica C* **1994,** *221*, 62.
[149] Y. Yokoyama, T. Kubo, Y. Nakagawa, M. Umeda, Y. Suzuki, S. Yoshida, *Physica C* **1994,** *219*, 327.
[150] L. A. Dorosinskii, M. V. Indenbom, V. I. Nikitenko, A. A. Polyanskii, R. L. Prozorov, V. K. Vlasko-Vlasov, *Physica C* **1993,** *206*, 360.
[151] I. K. Bdikin, I. M. Shmyt'ko, V. Sh. Shekhtman, N. V. Abrosimov, G. A. Emel'chenko, Yu. A. Ossipyan, *Physica C* **1992,** *201*, 69.
[152] K. Ogawa, J. Fujiwara, H. Takei, H. Asaoka, *Physica C* **1991,** *190*, 39.
[153] H. W. Zandbergen, *Physica C* **1992,** *194*, 287.
[154] W. Zhou, D. A. Jefferson, W. Y. Liang, *Superconductor Sci. Technol.* **1993,** *6*, 81.
[155] D. N. Davydov, A. Quivy, P. Diko, R. Deltour, M. Mehbod, M. Ye, *Solid State Commun.* **1993,** *86*, 267.
[156] H. L. Edwards, J. T. Markert, A. L. de Lozanne, *Phys. Rev. Lett.* **1992,** *69*, 2967.
[157] N. Schroeder, R. Böttner, S. Ratz, E. Dietz, U. Gerhardt, Th. Wolf, *Phys. Rev. B* **1993,** *47*, 5287.
[158] M. Lindroos, A. Bansil, K. Gofron, J. C. Campuzano, H. Ding, R. Liu, B. W. Veal, *Physica C* **1993,** *212*, 347.
[159] D. E. Fowler, C. R. Brundle, J. Lerczak, F. Holtzberg, *J. Electron Spectrosc. Related Phenom.* **1990,** *52*, 323.
[160] R. Liu, C. G. Olson, A.-B. Yang, C. Gu, D. W. Lynch, A. J. Arko, R. S. List, R. J. Barlett, B. W. Veal, J. Z. Liu, A. P. Paulikas, K. Vandervoort, *Phys. Rev. B* **1989,** *40*, 2650.
[161] N. Schroeder, S. Weiss, R. Böttner, S. Marquardt, S. Ratz, E. Dietz, U. Gerhardt, G. Ecke, H. Rössler, Th. Wolf, *Physica C* **1993,** *217*, 220.
[162] Udayan De, S. Natarajan, E. W. Seibt, *Physica C* **1991,** *183*, 83.
[163] M. Mokhtari, C. Perrin, O. Pena, M. Sergent, *Physica C* **1992,** *202*, 141.
[164] V. P. Dravid, Hong Zhang, L. D. Marks, J. P. Zhang, *Physica C* **1992,** *192*, 31.
[165] H. W. Zandbergen, R. Gronsky, K. Wang, G. Thomas, *Nature* **1988,** *331*, 596.
[166] E. Faulques, Ph. Molinié, P. Berdahl, T. Pap Nguyen, J.-L. Mansot, *Physica C* **1994,** *219*, 297.
[167] E. W. Muller in *Diffraction and Imaging Techniques in Materials Science* (Eds.: S. Amelinckx, R. Gevers, J. Van Landuyt), North Holland, Amsterdam **1978,** p. 791.
[168] S. I. Shkuratov, S. N. Ivanov, S. N. Shilimanov, *Physica C* **1993,** *213*, 321.
[169] I. M. Dmitrenko, P. A. Grib, A. G. Sivakov, O. G. Turutanov, A. P. Zhuravel, *Low Temp. Phys.* **1993,** *19*, 747.
[170] A. G. Sivakov, A. P. Zhuravel, O. G. Turutanov, I. M. Dmitrenko, J. W. M. Hilgenkamp, G. C. S. Brons, J. Flokstra, H. Rogalla, *Physica C* **1994,** *232*, 93.

[171] Yu. Ya. Divin, P. M. Shadrin, *Physica C* **1994,** *232*, 257.
[172] M. R. Koblischka, Th. Schuster, B. Ludescher, H. Kronmüller, *Physica C* **1992,** *190*, 557.
[173] L. A. Bursill, Xu Dong Fan, *Phys. Status Solidi* **1988,** *107*, 503.
[174] W. A. Zhou, D. A. Jefferenson, W. Y. Liang, *Surface Sci.* **1994,** *310*, 52.
[175] W. Zhou, D. A. Jefferson, W. Y. Liang, A. Soeta, T. Kamo, S. P. Matsuda, *Physica C* **1992,** *202*, 335.
[176] Y. Xin, W. Zhou, C. J. Humphreys, *Physica C* (in press).
[177] H. W. Zandbergen, P. Groen, G. Van Tendeloo, J. Van Landuyt, S. Amelinckx, *Solid State Commun.* **1988,** *66*, 397.
[178] H. W. Zandbergen, K. Kadowaki, M. J. V. Menken, A. A. Menovsky, G. Van Tendeloo, S. Amelinckx, *Physica* **1989,** *158*, 155.
[179] O. Eibl, B. Roas, *J. Mater. Res.* **1990,** *5*, 2620.
[180] C. Traeholt, J. G. Wen, H. W. Zandbergen, Y. Shen, J. W. M. Hilgenkamp, *Physica C* **1994,** *230*, 425.
[181] C. Rossel, A. Catana, R. R. Schulz, E. J. Williams, A. Perrin, M. Guilloux-Viry, C. Thivet, *Physica C* **1994,** *223*, 370.
[182] D. Hesse, L. Berthold, H. Haefke, H. P. Lang, R. Sum, H.-J. Güntherodt, *Physica C* **1992,** *202*, 277.
[183] P. Berberich, J. Tate, W. Dietsche, H. Kinder, *Appl. Phys. Lett.* **1988,** *53*, 925.
[184] H. P. Lang, H. Haefke, R. Sum, H.-J. Güntherodt, L. Berthold, D. Hesse, *Physica C* **1992,** *202*, 289.
[185] A. L. Vasiliev, G. Van Tendeloo, S. Amelinckx, Yu. Boikov, E. Olsson, Z. Ivanov, *Physica C* **1995,** *244*, 373.
[186] C. B. Eom, J. Z. Sun, B. M. Lairson, S. K. Streiffer, A. F. Marshall, K. Yamamoto, S. M. Anlage, J. C. Bravman, T. H. Geballe, *Physica C* **1990,** *171*, 354.
[187] K.-i. Shimura, Y. Daitoh, Y. Yano, T. Terashima, Y. Bando, Y. Matsuda, S. Komiyama, *Physica C* **1994,** *228*, 91.
[188] Z. X. Gao, I. Heyvaert, B. Wuyts, E. Osquiguil, C. Van Haesendonck, Y. Bruynseraede, *Appl. Phys. Lett.* **1994,** *65*, 770.
[189] N. Savvides, A. Katsaros, *Physica C* **1994,** *226*, 23.
[190] S. Takeno, S.-i. Nakamura, M. Sagoi, T. Miura, *Physica C* **1991,** *176*, 151.
[191] H. P. Lang, H. Haefke, G. Leemann, H.-J. Güntherodt, *Physica C* **1992,** *194*, 81.
[192] K. Verbist, A. L. Vasiliev, G. Van Tendeloo, *Appl. Phys. Lett.* **1995,** *66*, 1424.
[193] C. Traeholt, J. G. Wen, V. Svetchnikov, A. Delsing, H. W. Zandbergen, *Physica C* **1993,** *206*, 318.
[194] S. J. Klepeis, J. P. Benedict, R. Anderson, *EMSA proc.* (Ed: G. W. Bailey), San Francisco Press **1989,** p. 712.
[195] O. Eibl, H. E. Hoenig, J.-M. Triscone, Ø. Fischer, L. Antognazza, O. Brunner, *Physica C* **1990,** *172*, 365.
[196] C. Træholt, J. G. Wen, V. Svetchnikov, H. W. Zandbergen, *Physica C* **1994,** *230*, 297.
[197] L. A. Tietz, C. B. Carter, D. K. Lathrop, S. E. Rossik, R. A. Buhrman, J. R. Michael, *J. Mater. Res.* **1989,** *4*, 00.
[198] S. J. Pennycook, M. F. Crisholm, D. E. Jessorn, R. Feenstra, S. Zhu, X. Y. Zheng, D. J. Lowndes, *Physica C* **1992,** *202*, 1.
[199] J. G. Wen, C. Træholt, H. W. Zandbergen, K. Joosse, E. M. C. M. Reuvekamp, H. Rogalla, *Physica C* **1993,** *218*, 29.
[200] D. K. Fork, D. B. Fenner, R. W. Barton, J. M. Phillips, G. A. N. Connell, J. B. Boyece, T. H. Geballe, *Appl. Phys. Lett.* **1990,** *57*, 1161.
[201] J. Wecker, Th. Matthée, H. Benher, G. Friedl, K. Samwer, *Mater. Res. Soc. Symp. Proc.* **1992,** *275*, 1240.
[202] D. K. Fork, S. M. Garrison, T. H. Geballe, *J. Mater. Res.* **1992,** *7*, 1641.
[203] A. Bardal, O. Eibl, Th. Matthée, G. Friedl, J. Wecker, *J. Mater. Res.* **1993,** *8*, 2112.
[204] C. A. Copetti, H. Soltner, J. Schubert, W. Zander, O. Hollricher, Ch. Buchal, W. Schulz, N. Tellman, N. Klein, *Appl. Phys. Lett.* **1993,** *63*, 1429.
[205] Yu. Boikov, Z. G. Ivanov, A. L. Vasiliev, T. Claeson, *J. Appl. Phys.* **1995,** *77*, 1654.
[206] A. L. Vasiliev, G. Van Tendeloo, S. Amelinckx, Yu. Boikov, E. Olsson, Z. Ivanov, *Physica C* **1994,** *244*, 373.
[207] A. L. Vasiliev, G. Van Tendeloo, Yu. Boikov, E. Olsson, Z. Ivanov, *Phil. Mag.* (in press).
[208] W. Gerhäuser, H.-W. Neumüller, W. Schmidt, G. Ries, O. Eibl, G. Saemann-Ischenko, S. Klaumünzer, *Physica C* **1991,** *185–189*, 2339.
[209] W. Gerhäuser, G. Ries, H. W. Neumüller, W. Schmidt, O. Eibl, G. Saemann-Ischenko, S. Klaumünzer, *Phys. Rev. Lett.* **1992,** *68*, 879.
[210] W. L. Zhou, Y. Sasaki, Y. Ikuhara, *Physica C* **1994,** *234*, 323.
[211] H. Watanabe, B. Kabius, K. Urban, B. Roas, S. Klaumünzer, G. Saemann-Ischenko, *Physica C* **1991,** *179*, 75.
[212] V. Hardy, M. Hervieu, D. Groult, J. Provost, *Radiat. Effects Defects Solids* **1991,** *118*, 343.
[213] A. Wahl, M. Hervieu, G. Van Tendeloo, V. Hardy, J. Provost, D. Groult, Ch. Simon, B. Raveau, *Radiat. Effects* **1995,** *133*, 293.
[214] M. J. Kramer, Q. Qian, D. Finnemore, L. Snead, *Physica C* **1992,** *203*, 83.

[215] M. A. Kirk, H. W. Weber in *Studies of High Temperature Superconductors* (Ed.: A. V. Narlikar), Nova Science, New York **1992,** Vol. 10, p. 253.
[216] J.-W. Lee, H. S. Kessure, D. E. Laughlin, M. E. McHenry, S. G. Sankar, J. O. Willis, J. R. Cost, M. P. Maley, *Appl. Phys. Lett.* **1990,** *57*, 2150.
[217] M. C. Frischherz, M. A. Kirk, J. Farmer, L. R. Greenwood, H. W. Weber, *Physica C* **1994,** *232*, 309.
[218] H. W. Zandbergen, J. Kulik, B. Nieuwendijk, *Physica C* **1991,** *179*, 43.
[219] G. Van Tendeloo, M.-O. Ruault, H. Bernas, M. Gasgnier, *J. Mater. Res.* **1991,** *6*, 677.
[220] V. P. Dravid, H. Zhang, Y. Y. Wang, *Physica C* **1993,** *213*, 353.
[221] H. W. Zandbergen, W. T. Fu, K. Kadowaki, G. Van Tendeloo, *Physica C* **1989,** *161*, 390.
[222] H. W. Zandbergen, W. T. Fu, L. J. de Jong, G. Van Tendeloo, *Cryogenics* **1990,** *30*, 628.
[223] H. W. Zandbergen, R. Gronsky, G. Van Tendeloo, *J. Superconductivity* **1989,** *2*, 337.
[224] J. Y. Laval, W. Swiatnicki, *Physica C* **1994,** *221*, 11.
[225] O. Eibl, *Physica C* **1990,** *168*, 239.
[226] B. Kabius, J. W. Seo, T. Amrein, U. Dähne, A. Scholen, M. Siegel, K. Urban, L. Schultz, *Physica C* **1994,** *231*, 123.
[227] A. Umezawa, Y. Feng, H. S. Edelman, T. C. Willis, J. A. Parrell, D. C. Larbalestier, G. N. Riley Jr., W. L. Carter, *Physica C* **1994,** *219*, 378.
[228] K. B. Alexander, D. M. Kroeger, J. Bentley, J. Brynestad, *Physica C* **1991,** *180*, 337.
[229] N. D. Browning, M. F. Chisholm, S. J. Pennycook, D. P. Norton, D. H. Lowndes, *Physica C* **1993,** *212*, 185.
[230] R. Ramesh, S. Jin, P. March, *Nature* **1990,** *346*, 420.
[231] H. W. Zandbergen, *Physica C* **1992,** *193*, 371.
[232] T. Krekels, G. Van Tendeloo, S. Amelinckx, J. Karpinski, S. Rusiecki, E. Kaldis, E. Jilek, *Physica C* **1991,** *178*, 383.
[233] M. A. Senaris-Rodriguez, A. M. Chippindale, A. Varez, E. Moran, M. A. Alario-Franco, *Physica C* **1991,** *172*, 477.
[234] T. Siegrist, S. M. Zahurak, D. W. Murphy, R. S. Roth, *Nature* **1988,** *334*, 231.
[235] S. Tao, H.-U. Nissen, C. Beeli, M. Cantoni, M. G. Smith, J. Zhou, J. B. Goodenough, *Physica C* **1992,** *204*, 117.
[236] Z. Hiroi, M. Azuma, M. Takano, Y. Takeda, *Physica C* **1993,** *208*, 286.
[237] S. Adachi, H. Yamauchi, S. Tanaka, N. Môri, *Physica C* **1993,** *212*, 164.
[238] N. Ikeda, Z. Hiroi, M. Azuma, M. Takano, Y. Bando, Y. Takeda, *Physica C* **1992,** *210*, 367.
[239] N. Kijima, R. Gronsky, X.-D. Xiang, W. A. Vareka, A. Zettl, J. L. Corkill, M. L. Cohen, *Physica C* **1991,** *184*, 127.
[240] S. Nakajima, T. Oku, R. Suzuki, M. Kikuchi, K. Hiraga, Y. Syono, *Physica C* **1993,** *214*, 80.
[241] E. Ohshima, M. Kikuchi, F. Izumi, K. Hiraga, T. Oku, S. Nakajima, N. Ohnishi, Y. Morii, S. Funahashi, Y. Syono, *Physica C* **1994,** *221*, 261.
[242] T. Wada, A. Nara, A. Ichinose, H. Yamauchi, S. Tanaka, *Physica C* **1992,** *192*, 181.
[243] N. Kijima, R. Gronsky, X.-D. Xiang, W. A. Vareka, A. Zettl, J. L. Corkill, M. L. Cohen, *Physica C* **1992,** *190*, 597.
[244] N. Kijima, R. Gronsky, X.-D. Xiang, W. A. Vareka, J. Hou, A. Zettl, J. L. Corkill, M. L. Cohen, *Physica C* **1992,** *198*, 309.
[245] Y. Miyazaki, H. Yamane, N. Ohnishi, T. Kajitani, K. Hiraga, Y. Morii, S. Funahashi, T. Hirai, *Physica C* **1992,** *198*, 7.
[246] M. Isobe, T. Kawashima, K. Kosuda, Y. Matsui, E. Takayama-Muromachi, *Physica C* **1994,** *234*, 120.
[247] J. Akimitsu, M. Vehara, M. Ogawa, H. Nakata, K. Tomimoto, Y. Myyazaki, H. Yamane, T. Hirai, K. Kinoshita, Y. Matsui, *Physica C* **1992,** *201*, 320.
[248] A. Maignan, M. Hervieu, C. Michel, B. Raveau, *Physica C* **1993,** *208*, 116.
[249] K. Kinoshita, T. Yamada, *Nature* (*London*), **1992,** *337*, 312.
[250] F. Izumi, K. Kinoshita, Y. Matsui, K. Yanagisawa, T. Ishigaki, T. Kamiyama, T. Yamada, H. Asano, *Physica C* **1992,** *196*, 227.
[251] M. Huvé, C. Michel, A. Maignan, M. Hervieu, C. Martin, B. Raveau, *Physica C* **1993,** *205*, 219.
[252] A. Maignan, M. Huvé, C. Michel, M. Hervieu, C. Martin, B. Raveau, *Physica C* **1993,** *208*, 149.
[253] D. Pelloquin, M. Caldès, A. Maignan, C. Michel, M. Hervieu, B. Raveau, *Physica C* **1993,** *212*, 199.
[254] D. Pelloquin, A. Maignan, M. Caldès, M. Hervieu, C. Michel, B. Raveau, *Physica C* **1993,** *212*, 199.
[255] M. H. Pan, M. Greenblatt, *Physica C* **1991,** *176*, 80.
[256] M. H. Pan, M. Greenblatt, *Physica C* **1991,** *184*, 235.
[257] C. Michel, M. Hervieu, M. M. Borel, A. Grandin, F. Deslandes, J. Provost, B. Raveau, *Z. Phys. B* **1987,** *68*, 421.
[258] F. Goutenoire, M. Hervieu, C. Michel, C. Martin, B. Raveau, *Physica C* **1993,** *210*, 359.
[259] M. Uehara, S. Sahoda, H. Nakata, J. Akemitsu, Y. Matsui, *Physica C* **1994,** *222*, 27.

[260] M. Huvé, G. Van Tendeloo, M. Hervieu, A. Maignan, B. Raveau, *Physica C* **1994,** *231*, 15.
[261] M. Huvé, G. Van Tendeloo, S. Amelinckx, M. Hervieu, B. Raveau, *J. Solid State Chem.* **1995,** *114*, 560.
[262] S. N. Putilin, E. V. Antipov, O. Chmaissen, M. Marezio, *Nature (London)*, **1993,** *362*, 226.
[263] D. Pelloquin, C. Michel, G. Van Tendeloo, A. Maignan, M. Hervieu, B. Raveau, *Physica C* **1993,** *124*, 87.
[264] A. Schilling, M. Cantoni, J. D. Guo, R. Ott, *Nature (London)*, **1993,** *363*, 56.
[265] S. N. Putilin, E. V. Antipov, M. Marezio, *Physica C* **1993,** *212*, 266.
[266] A. Maignan, G. Van Tendeloo, M. Hervieu, C. Michel, B. Raveau, *Physica C* **1993,** *212*, 239.
[267] A. Maignan, C. Michel, G. Van Tendeloo, M. Hervieu, B. Raveau, *Physica C* **1993,** *216*, 1.
[268] S. F. Hu, D. A. Jefferson, R. S. Liu, P. P. Edwards, *J. Solid State Chem.* **1993,** *103*, 280.
[269] D. Pelloquin, M. Hervieu, C. Michel, G. Van Tendeloo, A. Maignan, B. Raveau, *Physica C* **1993,** *216*, 257.
[270] C. Martin, D. Bourgault, C. Michel, J. Provost, M. Hervieu, B. Raveau, *Eur. J. Solid. State Inorg. Chem.* **1989,** *26*, 1.
[271] F. Goutenoire, P. Daniel, M. Hervieu, G. Van Tendeloo, C. Michel, A. Maignan, B. Raveau, *Physica C* **1993,** *216*, 243.
[272] C. Martin, M. Huvé, G. Van Tendeloo, A. Maignan, C. Michel, M. Hervieu, B. Raveau, *Physica C* **1993,** *212*, 274.
[273] A. Maignan, C. Martin, M. Huvé, J. Provost, M. Hervieu, C. C. Michel, B. Raveau, *Physica C* **1990,** *170*, 350.
[274] M. Marezio, E. V. Antipov, J. J. Capponi, C. Chaillout, S. Loureiro, S. N. Putilin, A. Santoro, J. L. Tholence, *Physica B* **1994,** *197*, 570.
[275] E. V. Antipov, S. M. Loureiro, C. Chaillout, J. J. Capponi, P. Bordet, J. L. Tholence, S. N. Putilin, M. Marezio, *Physica C* **1993,** *215*, 1.
[276] S. Nakajima, M. Kikuchi, Y. Syono, T. Oku, D. Shindo, K. Hiraga, N. Kobayashi, H. Iwasaki, Y. Muto, *Physica C* **1989,** *158*, 471.
[277] M. Cantoni, A. Schilling, H.-U. Nissen, H. R. Ott, *Physica C* **1993,** *215*, 11.
[278] G. Van Tendeloo, C. Chaillout, J. J. Capponi, M. Marezio, E. V. Antipov, *Physica C* **1994,** *223*, 219.
[279] C. Martin, M. Hervieu, G. Van Tendeloo, F. Goutenoire, C. Michel, A. Maignan, B. Raveau, *Solid State Commun.* **1995,** *93*, 53.
[280] E. M. Kopnin, E. V. Antipov, J. J. Capponi, P. Bordet, C. Chaillout, S. de Brion, M. Marezio, A. P. Bobylev, G. Van Tendeloo, *Physica C* **1995,** *243*, 222.

Handbook of Microscopy
S. Amelinckx, D. van Dyck, J. van Landuyt, G. van Tendeloo

9 Non-Periodic Structures

9.1 High-Resolution Imaging of Amorphous Materials

9.1.1 Introduction

It is tempting to consider the atomic structure of an amorphous solid as a random collection of atoms, but for most glasses and amorphous thin films this is far from accurate. Both positional and compositional order are relatively well-defined over short distances—nearest and next nearest atomic neighbors—what is often called the 'local structure'. It is possible to define local structural units such as $SiSi_4$ or SiO_4 tetrahedra in amorphous Si or glassy SiO_2. These have symmetry and geometry equivalent to corresponding crystalline phases and the way in which the units are connected is often recognizably the same as that of the crystal, although deviations from regularity are much more pronounced. Beyond about 2 nm the long-range structure of all amorphous solids can be represented by truly averaged properties (except for compositional inhomogeneities in materials which exhibit immiscibility).

In contrast, the 'medium-range structure'—from about 0.6 to 2 nm—is difficult to determine and to specify [1]. Diffraction techniques, which have provided most of the structural information on amorphous solids, become decreasingly useful beyond third neighbors, for obvious reasons. X-ray and neutron diffraction involve massive averaging, over at least 10^{20} atoms, over all types of pair correlations $P_{\alpha\beta}(r)$ in polyatomic solids, and over all directions in space. The distribution functions are one-dimensional (1-D) and the various functions $P_{\alpha\beta}(r)$ overlap—usually in antiphase due to charge ordering—thus submerging information about the spatial (or temporal) variability of the structure. Any structural ordering revealed by diffraction thus represents a lower bound—the actual structure may be much more organized (and more exciting).

The paucity of structural information beyond about 0.6 nm obtainable by diffraction techniques can be complemented by high resolution electron microscopy (HREM). Ideally, the amplitude and phase of the scattered electron wave is retained in the imaging process so that structural information from a microscopic region of a specimen is retained in a real space HREM image, albeit in a scrambled form. Averaging is thus reduced by many orders of magnitude. The effective specimen volume is now typically a 4 nm cube containing, say, 10^4 atoms, and the image is a two-dimensional (2-D) rather than a 1-D

representation of the three-dimensional (3-D) structure.

Studies of the structure of amorphous solids by electron microscopy have, however, produced very variable results, often depending on the type of material and on the questions asked. Glasses are metastable with respect to the crystalline state and the detection of nucleation and incipient crystallization represents one of the simplest problems. Both bright-field and dark-field medium-resolution electron microscopy give unambiguous identification of crystallites that are (typically) larger than 2 nm in diameter. Similarly, the presence of voids or columnar structure on a scale of a few nanometers can be tackled with considerable confidence. A single-phase polyatomic glass may also be metastable with respect to phase separation into two or more immiscible supercooled liquid phases of different compositions, on a scale of a few nanometers upwards. Examination of the detailed morphology, denoting spinodal or 'nucleation and growth mechanisms' represents one of the major triumphs of electron microscopy applied to this field [2].

However, this chapter concentrates on the more open (and demanding) questions surrounding the medium-range structure of amorphous solids studied by high resolution microscopy.

HREM bright-field images of glasses and amorphous semiconductors usually display a grainy structure on a scale of 0.2–0.4 nm. This is sometimes described as a 'pepper and salt' structure, although it is usually possible to recognize 'fringes' that may be parallel and sometimes straight over areas of 2–5 nm^2. The appearance is similar to (distorted) lattice fringes of crystalline solids and, as there may be some regularity and symmetry, and the fringe spacings are often characteristic of similar crystalline phases, the possibility arises of treating these as coherently scattering domains. Similar effects may be seen in tilted illumination dark-field images, where the coherently diffracting regions appear as bright patches in an otherwise random speckle pattern. Here we concentrate on bright-field HREM images.

Axial bright-field images of several amorphous alloys show pronounced fringe-like detail on a scale of 1–2 nm: results that can be linked to the close-packed structures of these glasses. For the more open structures typical of elemental amorphous semiconductors and oxide glasses, the situation is less clear. Apart from a few cases (e.g. near an interface with a crystal or in very thin specimens), the amount of structural information is very limited. To date, HREM has provided few really essential clues to the nature of the amorphous state. But the situation is changing rapidly, and it is perhaps not overoptimistic to claim that within 10 years the last statement will be redundant.

Optimism is based on two recent advances. Firstly, improvements in resolution limits now allow state-of-the-art microscopes to image, routinely, atomic density fluctuations in the region of 0.2 nm (typical of the detail seen in amorphous metals), and even more adequately to record features of the structure of oxide glasses and amorphous elemental semiconductors. Secondly, it is now possible to compare experimental HREM images with simulated images of atomic models that are sufficiently large to give realistic results. The process is aided, of course, by the marketing of user-friendly software

packages for computing images and to modify atomic models.

9.1.2 Obstacles to Progress

9.1.2.1 Resolution Limit

A fundamental problem that has colored the history of HREM investigations of amorphous solids stems from the fact that the structural information is delocalized in reciprocal space. Instead of the narrow diffraction rings typical of a polycrystalline solid, scattering from amorphous solids gives a series of diffuse halos with rapidly decreasing intensity as the scattering angle increases. Neglecting scattering from the surfaces, inelastic scattering etc., the orientationally averaged scattered intensity from a specimen containing N atoms can be represented by the static structure factor:

$$S(Q) = \frac{\left|\sum_{i=1}^{N} \exp(i\boldsymbol{Q}\cdot\boldsymbol{R}_i)\right|^2}{N} \qquad (1)$$

where $\boldsymbol{R}_i$ is the position vector of the ith atom, $\boldsymbol{Q}$ is the scattering vector, and $Q = |\boldsymbol{Q}| = 4\pi \sin\theta/\lambda$; 2θ and λ are the scattering angle and electron wavelength. For previous generations of microscopes the information limit extended to (at best) only the first peak of $S(Q)$ at $Q_1 = 20$ to $30\,\mathrm{nm}^{-1}$, including one oscillation of the contrast transfer function. Attempts were made to extend the range of Q-space sampled by using tilted illumination, but the benefits this offers are outweighed by the image artifacts introduced.

In terms of the Abbe theory, the electron wave scattered by the specimen is

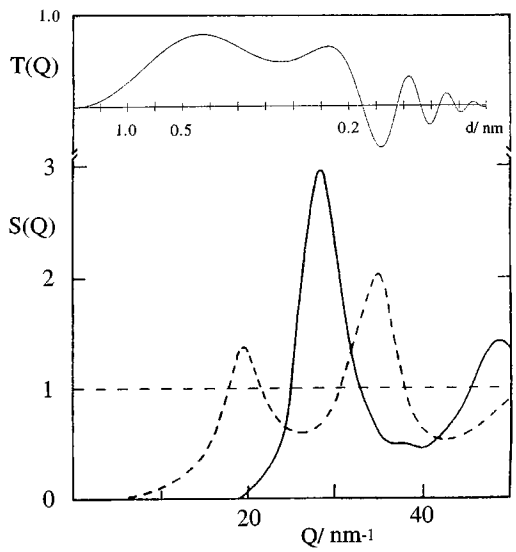


Figure 1. The contrast transfer function $T(Q)$ for a JEM200CX with a high-resolution polepiece [3] at a defocus of 45 nm, together with the structure factor $S(Q)$ for amorphous Ge and amorphous Pd_4Si. Interplanar spacings ($d = 2\pi/Q$) are marked in the upper diagram.

modified in the diffraction plane (the back focal plane of the objective lens) by a scattering angle-dependent instrument response function: the contrast transfer function (CTF), denoted here by $T(Q)$. Typical $S(Q)$ data and $T(Q)$ for a JEM200CX with a specially designed high resolution polepiece which allows a point-to-point resolution limit of slightly less than 0.2 nm [3] is shown in Fig. 1. Structural information is contained in the oscillations of $S(Q)$ and the information transferred to the real-space image corresponds to regions of Q-space where the scattered amplitude,

$$\sum_{i=1}^{N} \exp(i\boldsymbol{Q}\cdot\boldsymbol{R}_i)$$

and, therefore, $S(Q)$, departs from unity and $T(Q)$ is nonzero. Negative and

positive values of $T(Q)$ imply contrast reversal—bright features in the image would become dark if $T(Q)$ changed sign, and information is suppressed near zeros of $T(Q)$. For faithful image reproduction, all the oscillations of $S(Q)$ should lie within the first lobe of $T(Q)$, which is clearly impossible (see Fig. 1). Even for current high-voltage microscopes, only the first peak of $S(Q)$ for most amorphous alloys lies within the first lobe. The problem is less severe for crystals—diffraction space information is localized at a few discrete peaks, and it is often possible to arrange suitable values of $T(Q)$ at these points to give adequate imaging.

9.1.2.2 The Projection Problem

A second major problem is overlap of the images along a column normal parallel to the beam direction z. For a thick aperiodic specimen, the projected potential of a column of atoms parallel to z corresponds to dense, randomly positioned cluster of points in the x,y plane [4, 5]. As the medium-range structure under examination is unlikely to exceed 2 nm in any direction, the image of a 10 nm thick film consists of overlapping contributions from about 4–5 such regions. Atomic detail is thus undecipherable in specimens thicker than about 0.5–1 nm, and foils thicker than about 10 nm can be treated as white (or strictly grey) noise objects. Calculations by Bursill et al. [6] show this to be the case: reliable atomic positions are obtainable only for films thinner than about 0.6 nm. Most HREM work has assumed this result and has been directed to the simpler task of assessing the extent of positionally ordered regions, with characteristics similar to microcrystallites or aperiodic 'amorphous clusters'.

This appears to be a most useful area for detailed study, with the likelihood that HREM measurements can provide significant, experimental structural evidence to complement X-ray and neutron scattering data on the local structure. There is evidence that even in 'random' atomic models for glasses, and particularly metallic glasses, 'pseudo-Bragg' features are detectable in projections of the atomic coordinates of the model. Early work on this topic by Alben et al. [7] has been extended by Wallis [8], Gaskell and Wallis [9], and by Mountjoy [10]. These latter authors have also related the quasi-Bragg planes to the position of the first (sharp) diffraction peak seen in silicates and many other amorphous solids for which no generally agreed explanation exists.

Recognizable (Bragg-orientated) structure in an HREM image occupies only a fraction of the volume and the remainder (randomly positioned atoms or Bragg 'misorientated' regions) generate noise. Krivanek [11] has given a useful rule of thumb to estimate the minimum size of an orientated region visible in an otherwise disordered foil—planes of Ge in a microcrystallite will be visible if the diameter is greater than about one-third of the foil thickness. Calculations were for the specific case of {111} planes of the diamond-cubic structure which are not densely populated, so that the estimate should provide a reasonable lower bound for many materials. This estimate has been confirmed by image simulation of an approximately 0.5 nm crystallite of γ-Pd_3Si in a matrix of dense random-packed alloys of the same composition [12] and

similar work [13] on amorphized regions in γ-$ZrSiO_4$. Crystalline lattice fringes are almost impossible to detect for crystal/total foil thicknesses less than about 0.23. Interestingly, the corresponding diffraction spots are identifiable in the optical diffractogram at a foil thickness of about twice this value.

9.1.2.3 Electron Beam Damage and Similar Specimen-Dependent Problems

Those used to the art of recording high-quality HREM images of amorphous solids will recognize the need for careful alignment of the optics. Clearly, correction of astigmatism, current and voltage centering [14] are essential at the time of taking a through-focal series of images. Examination of the images after the experiment is also vital, to check for absence of instrumental aberrations, specimen drift etc., and to select the most appropriate defocus conditions. In this regard the two-dimensional Fourier transform of the image—that is, the power spectrum, obtained by optical diffraction or computer processing—can be an essential diagnostic tool. The power spectrum should consist of continuous, circular bright rings with dark rings corresponding to zeros of the CTF [14]. During the experimental phase, such careful alignment takes time. There may also be the need for high current density, simply to see image defects, so that electron-beam damage may irrevocably alter the specimen. Even adopting the stratagem of aligning the optics on one part of the specimen, then recording images on another—perhaps even 'shooting blind' and relying on post-experiment inspection to select appropriate areas, defocus conditions etc.—may prove fruitless, to the extent that many materials become totally inaccessible to HREM studies. Oxide glasses containing mobile cations such as Na^+ or Li^+, low-melting chalcogenide glasses, and even some relatively refractory amorphous metals can be difficult or impossible to study. Crystallization under the beam is, of course, an additional problem and diffraction patterns recorded at the beginning and end of a through-focal series are needed to examine this possibility.

The general rule of thumb is that glasses should be refractory and have a low ionic conductivity—like silica, say. Amorphous alloys may have the advantage of being metals and having a reasonable electronic conductivity, thereby reducing beam-induced damage. But they also need to be resistant to oxidation or sublimation. For materials that are prone to oxidation (e.g. Fe- or Ti-containing alloys) 1 nm of the oxide coating each surface of a 4 nm specimen can ruin image interpretation as well as adding to the noise, so that the use of in situ techniques (e.g. Sinclair et al. [15]) or specially built devices to transfer the specimen from the preparation chamber may be inevitable.

9.1.3 Information Content of Axial Bright-Field Images

Interwoven with attempts to obtain convincing electron micrographs of amorphous solids have been the theoretical and computational studies of the information content of HREM images of amorphous solids. The subject has been reviewed in depth by

Howie [4, 5]; the more recent work will be emphasized here.

Problems arise partly because the image approximates to a projection of the atomic density through the thickness of the foil. While information is localized in the x, y plane of the specimen, correlations observed in a micrograph can arise from the chance correspondence of atoms separated by large distances in the z direction. It is thus necessary to judge whether the appearance of a micrograph represents more order than that associated with random arrangements of atoms or local structural groups. Problems are compounded by aberrations of the microscope. Even if the power spectrum of the specimen is featureless, corresponding to a white-noise object, oscillations in the contrast transfer function can produce an image which in previous generations of electron microscopes, can appear remarkably similar to an experimental image [16]. It is thus critical to compare the observed image with that expected for the material without the structural features being sought but, preferably, with those – such as local structural features – that can be assumed or are known from diffraction studies.

The least satisfactory approach – apart from computational convenience – is to use a completely random object as the model. The problem is that when an experimental micrograph is shown to contain features that are nonrandom, as in the analysis of images of α-Si recorded with a 400 kV instrument [17], where the question arises of the extent to which this reflects the known local ordering in α-Si, extending to (at least) second and third neighbors. As indicated above, perhaps the most satisfactory method is to simulate the image from atomic coordinates for a realistic atomic model that gives a good representation of the diffraction data. Alben et al. [7] discussed the order found in models for monoatomic amorphous metals and tetrahedral semiconductors. More convincingly, Schmidt et al. [18] have simulated bright-field images, including instrumental aberrations, of models for α-SiO_2, and have compared them with their experimental data.

An alternative method is to randomize the phases of the Fourier components which comprise the image, while preserving the amplitude (and thus the power spectrum). As mentioned above, the advantage offered by HREM over scattering techniques is preservation of the amplitude and phase of the scattered waves. The nonrandom character of the phase thus represents a key diagnostic. If the phase were a random function of the x, y coordinates, then positional order must be nonexistent. By randomizing the phase, by optical filtering [19] or computer simulation [17, 20, 21], a 'zero information' image can be created with the same power spectrum as the original. If the original, experimental micrograph contains structure additional to that seen in the zero information image, then that structure can be considered to be real—and otherwise, not. An approximate (but very convenient) method involves comparison of images from thin (<5 nm) and thick (10–15 nm) regions of the same specimen. Both have approximately the same power spectrum and are observed under almost identical microscope conditions, but only the thin specimens can be expected to show structurally significant features. The thicker regions thus provide their own internal standard.

Another technique for detecting ordered regions submerged by 'noise'

involves two-dimensional (2-D) correlation techniques [21–23]. The principle is to form the 2-D correlation function (in one dimension, the Patterson function, equivalent to the pair correlation function), which for an arbitrary function $f(\boldsymbol{r})$ is given by:

$$P(\boldsymbol{r}) = \int_{\boldsymbol{R}} f(\boldsymbol{R}) f(\boldsymbol{r}+\boldsymbol{R})\,\mathrm{d}\boldsymbol{R} \qquad (2)$$

$P(\boldsymbol{r})$ is proportional to the probability of finding an atom at $\boldsymbol{r}$ relative to any given atom site. This is applied to the 2-D intensity distribution in the image. The method, as described by Timsit et al. [22, 23] considers a microcrystallite of Si in a matrix of (simulated) α-Si. The central region of the image of the crystal was used as a 'template' to define the pixels representing the vectors $\boldsymbol{r}$. Specifically, the correlation function is (apart from normalization terms):

$$P(m,n) = \sum_{j=1}^{J} \sum_{k=1}^{K} I(j,k) I(j+m, k+n) \qquad (3)$$

where $P(m,n)$ is the value of the cross-correlation function at pixel (m,n), $I(p,q)$ is the image intensity at pixel (p,q) within the $P \times Q$ image area, and (j,k) are indices within the $J \times K$ template area. The image region occupied by the crystal which corresponds to the intensity pattern of the crystal template is highlighted in $P(m,n)$ and Timsit et al. showed that a template of one-third to half the crystal dimension is needed to give convincing results. The effectiveness of this process depends, of course, on a good choice of the symmetry of the template, which for an unknown specimen would involve extensive trial and error.

Attempts have been made by Fan and Cowley [17] to introduce another quantitative and objective measure of order. They show that departures from randomness can be expressed in terms of a relative entropy, defined in terms of the deviation of the intensity distribution in a micrograph from the most probable (gaussian) distribution. Specifically, the relative entropy Q_e is defined by:

$$Q_e = -N \sum f_j \ln\left(\frac{p_j}{f_j}\right) \qquad (4)$$

where f_j is the probability that an image pixel has intensity I_j (where j is one of n grey-scale values), p_j is the probability given by the entropy-maximizing Gaussian distribution, and N is the number of image pixels. Fan and Cowley show that $Q_e/n = 0.5$ for a simulated random image of a-Si, but for an experimental image (which shows very few signs of order by visual inspection) Q_e/n is 1.9, a significant departure from 0.5. The relative entropy method has also been used to gauge the information content of different members of a through-focal series and to assess the value of phase randomization techniques.

In the following sections, an attempt is made to survey the information that has been obtained, despite the problems mentioned, in a range of amorphous solids.

9.1.4 Amorphous Metals

A number of studies of amorphous metals by HREM have now been published and the general conclusion is that, providing the microscope has adequate point-to-point resolution to image the close-packed planes (~0.23 nm) and the foils are sufficiently thin (2–6 nm), then a surprising degree of positional ordering is observed.

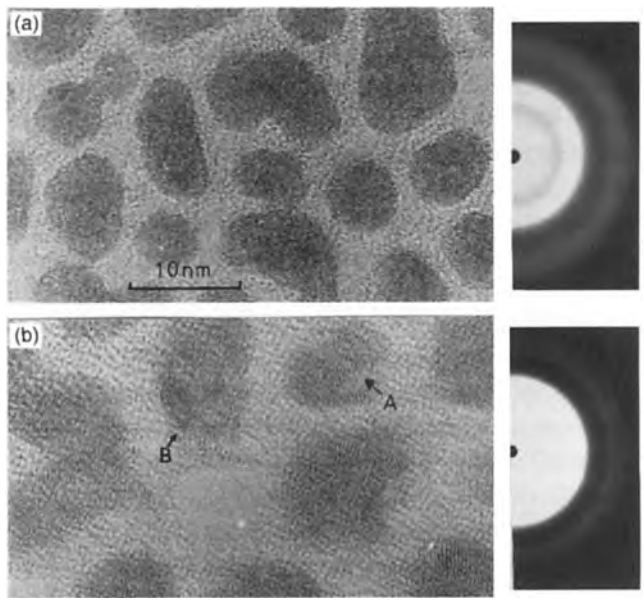


Figure 2. (a) High-resolution image of a particulate Pd–Si alloy recorded with axial illumination at 500 kV. The diffraction pattern is shown on the left. (b) Image of an α-Pd–Si alloy showing increased signs of order (as indicated by a splitting of the second diffraction peak). Note the resemblance to twins in the particles marked A and B. (Adapted from Gaskell and Smith [25].)

The early work of Gaskell, Smith and coworkers [24, 25] showed that thin discontinuous clusters of α-Pd–Si alloys prepared by flash evaporation onto a cooled rocksalt substrate exhibited 'lattice fringes' covering areas as large as 4–6 nm^2 in thin regions of the specimen (Fig. 2a). The general appearance of small distorted domains, similar to those seen in (larger) multiply twinned particles, is enhanced if the evaporation rate is reduced—twin boundaries can now be seen clearly (Fig. 2b). The diffraction pattern also shows some increased order.

More recent work by Hirotsu and coworkers [26, 27] shows that an even more stable alloy, α-Pd–Cu–Si, exhibits pronounced positional ordering, with parallel, often straight, lattice fringes over areas of about 4 nm^2. Again, the degree of order is a function of preparation conditions or of annealing. A typically careful reinvestigation of Pd–Si alloys by Anazawa et al. [28] serves as an example of this type of investigation. Amorphous thin films of $Pd_{82}Si_{18}$ were prepared by sputtering and the composition subsequently analyzed. Images were recorded in a 200-kV, high-resolution microscope with an interpretable resolution limit of less than 0.2 nm. Part of a focal series of the as-prepared film (Fig. 3) shows the presence of ordered regions, with well-defined fringes of spacing 0.19 and 0.22 nm with some geometrical symmetry. These resemble images simulated by a multislice method from a model of the same composition but with an embedded face-centered cubic (f.c.c.) cluster (Fig. 4). The electron diffraction data show classically diffuse halos typical of amorphous alloys, but the HREM images color this interpretation and give clear hints of the nature and extent of the medium-range ordering. Films of the same composition annealed at temperatures up to 673 K show the development of crystallites, with associated changes in the diffraction pattern.

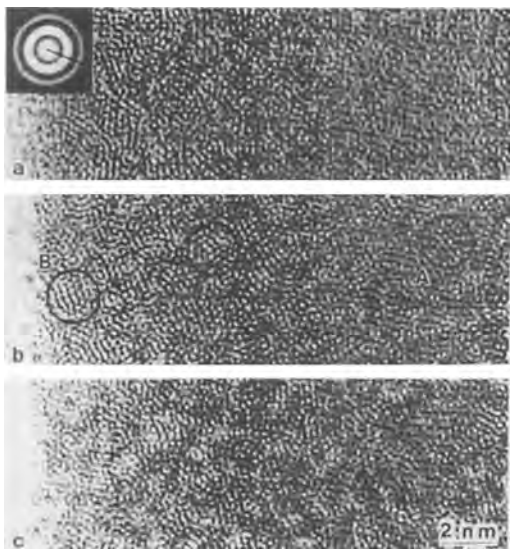


Figure 3. HREM images of α-$Pd_{82}Si_{18}$ as a function of defocus: (a) 40 nm; (b) 80 nm; (c) 120 nm. The electron diffraction pattern (upper left) displays classical diffuse halos. (Courtesy of Anazawa et al. [28].)

The images shown in Figs. 2 and 3 pose the question of how order in amorphous solids should be characterized. Except where stated, the diffraction patterns indicate no serious departures from the characteristic diffuse 'amorphous' rings, but both bright- and dark-field images suggest a degree of microcrystallinity. Dubois et al. [29] have shown that correlated, positionally ordered domains are not inconsistent with the diffraction data for α-Ni–B alloys, and there is no reason to suppose that this is a special case. Moreover, simulation of images of atomic models built with no impressed (or inserted) elements of translational order do, nevertheless, display features similar to the experimental images of the as-prepared glasses. Illustrations of this point are provided by simulated images for several models of

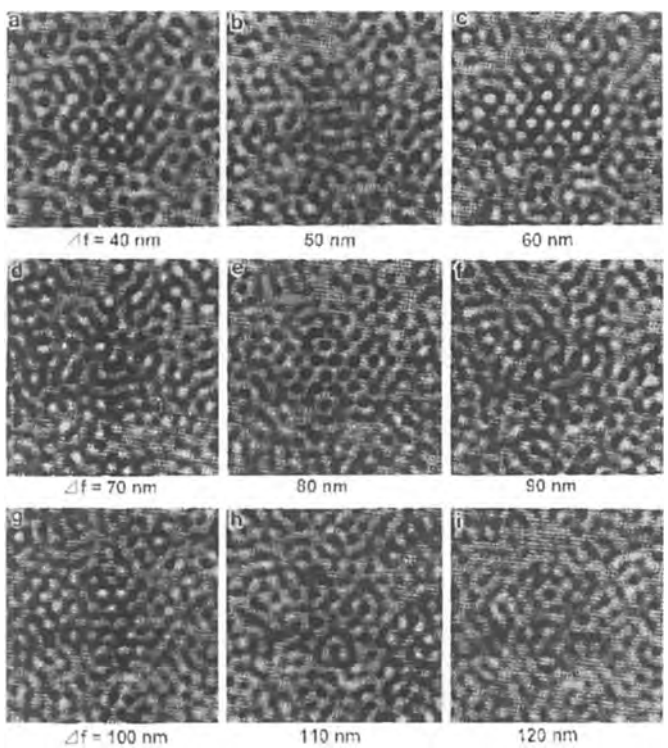


Figure 4. Simulated HREM images of an atomic model of $Pd_{82}Si_{18}$ at different defocus values Δf. The model consists of a random matrix of the alloy with an f.c.c. crystallite of the same composition embedded in the center. (Courtesy of Anazawa et al. [28].)

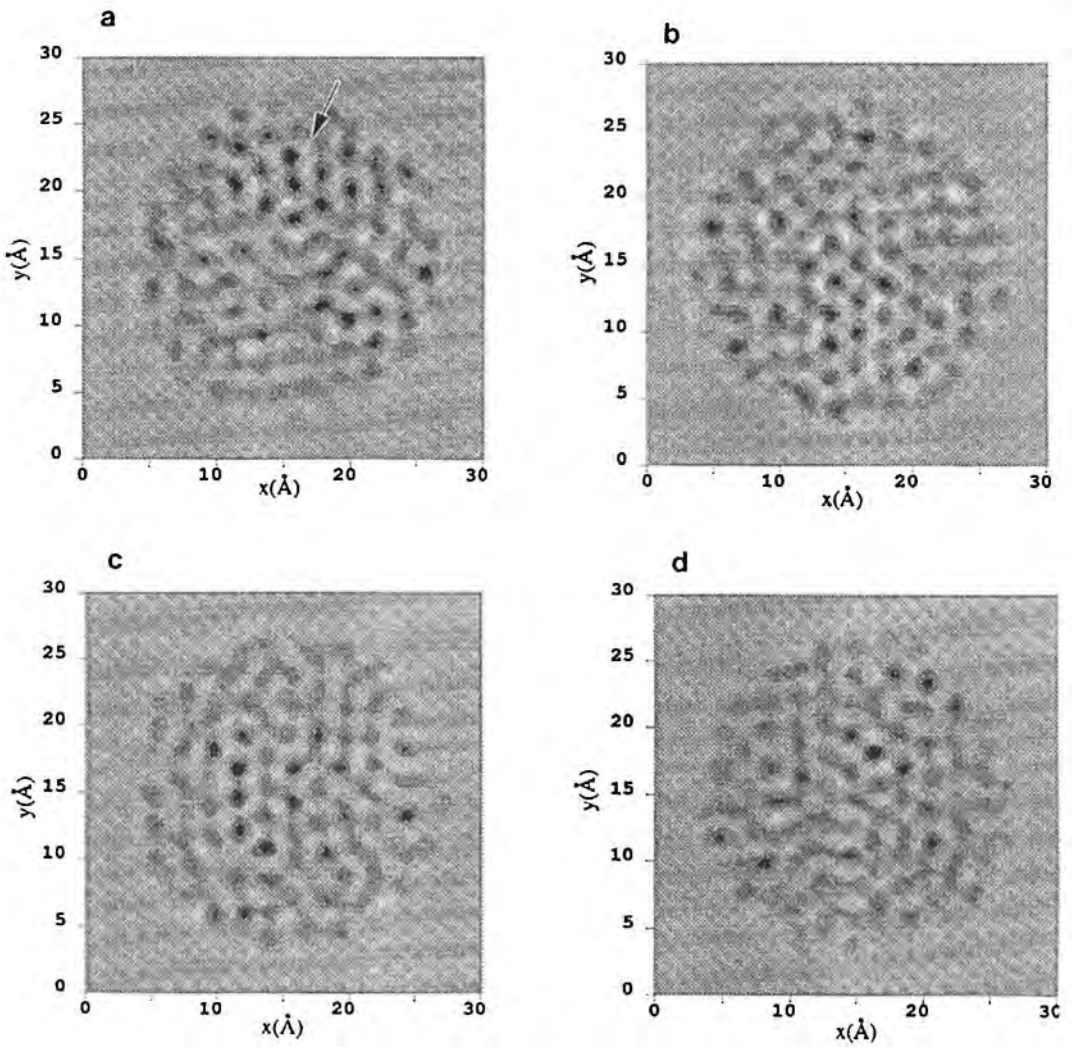


Figure 5. Simulated HREM images of several atomic models for amorphous monoatomic metals (or alloys). In each case, the models are orientated so that the scattering vectors corresponding to the two highest values of the anisotropic scattered intensity are in the plane of the diagram and (i) horizontal and (ii) at about 1 o'clock. Planar fluctuations in atomic density are thus viewed end-on and either vertically or at 10 o'clock. (a) A 500-atom spherical core cut from the center of Finney's 7994-atom model [30]; (b) a similar segment nearer the edge of the same model; (c) 1659-atom model consisting of the Ni coordinates from Beyer and Hoheisel's [31] model for α-$Ni_{81}B_{19}$; (d) 1456-atom model with the metal atom positions from Dubois et al.'s model [29] for binary metallic alloys of composition Pd_4Si, etc.

amorphous monatomic metals or alloys (Fig. 5) [10].

There is little reason, therefore, other than prejudice, to regard materials like a-$Pd_{82}Si_{18}$ as being somewhat less than ideal glasses and the detail of the medium-range structure offered by this type of investigation is challenging and exciting. It should be possible to establish whether the order, revealed by neutron scattering principally,

extending to about 1 nm in several typical amorphous alloys, is essentially microcrystalline, multiply twinned, or simply relates to the atomic density oscillations (as in Fig. 4), as expected in essentially random, densely packed solids subject to volume constraints. The existing HREM images provide some support for each of these possibilities, without definitely eliminating any.

9.1.5 Oxides

Amorphous solids can be formed from open structures typified by a-SiO_2 to close-packed oxides such as MoO_3 or WO_3. Oxides possess the distinct advantage for HREM work that the characteristic lattice spacings are larger than in metals, so that even 100 kV instruments can produce acceptable results, but most open-packed structures display less prominent image features. Images of the close-packed amorphous solids show more tendency toward medium-range ordering; for example, both MoO_3 and WO_3 show relatively well-defined, small ordered domains. WO_3 is a particularly interesting case because as-prepared specimens, heat-treated at relatively low temperatures, transform progressively and continuously to a microcrystalline state [31–33].

In SiO_2, however, correlated regions of lattice fringes are seen only rarely, in selected cases. For example, small clusters (about 10 nm in diameter) precipitated from aqueous solution were shown, by examination of the bright-field image and optical microdiffraction patterns, to be more ordered than similar material condensed from a high temperature, gas-phase reaction [34]. Bursill and Thomas [35] conclude that a commercial silica gel shows no evidence of nonrandom features.

9.1.6 Amorphous Tetrahedral Semiconductors

More effort has been devoted to imaging amorphous semiconductors and to attempts to interpret them, than has been spent on all other amorphous solids combined. In some ways this is a pity. Despite their commercial potential, amorphous elemental and compound semiconductors are not typical of amorphous solids. They are not glasses (the glass transition is hidden by crystallization), as they cannot be obtained from the melt except by ultra-rapid quenching. Moreover, thin films are prone to oxidation, and the first peak of the structure factor is weak. Comprehensive reviews of the subject have been given by Howie [4, 5].

Interesting images of α-Si and Ge were presented by Smith et al. [36] who observed little nonrandom structural detail in films around 10 nm thick, as expected [11]. Thinner films of α-Ge, notionally 1 nm thick, are relatively rich in nonrandom features and this finding is also supported by phase-randomization studies.

Some controversial images of α-Si deposited on a cleaved γ-Si wafer in UHV were reported by Phillips et al. [37] and Ourmazd et al. [38]. A cross-section near the interface between the crystalline Si and the amorphous film was examined using a 400-keV instrument, revealing ordered regions that were claimed to be 3 nm 'submicrocrystallites' on the evidence

of lattice fringes and optical microdiffraction patterns. The diffraction spots were small and orientationally related to the planes of the substrate. The extent of ordering, both positional and orientational, decreased with distance from the interface up to about 80 nm. The term 'orientational proximity effect' (OPE) was coined to describe the phenomenon. However, the generality of the OPE has yet to be confirmed, and there has been some difficulty in reproducing the effect under different conditions. The presence of ordered regions in amorphous solids near an interface with a compositionally similar crystal has been looked for on other occasions and not seen. Gaskell and Saeed [39] have attempted to observe the effect in Ge by sequentially depositing thin films of α-Ge over small islands of γ-Ge, and have examined the interfacial regions. While there is some evidence of order in the very thin regions—the images obtained were similar to those found by Smith et al. [36]—the fringes are not preferentially located near the interface and have no orientational relationship to the planes of the crystal. The effect may be limited – to certain types of interface, or to specific elements – or it might be destroyed by a thin oxide film that forms in vacuum (not UHV) between sequential depositions of γ- and α-Ge. However, there are instances where cross-sections of interfaces show some orientational effect, even when the two materials joined at the interface are different (e.g. the Si–SiO_2 interface [40]), although other studies [13] have shown that it is not always obvious just where the interface lies between a crystal and an amorphous overlayer, making it difficult to assess the spatial extent of the OPE.

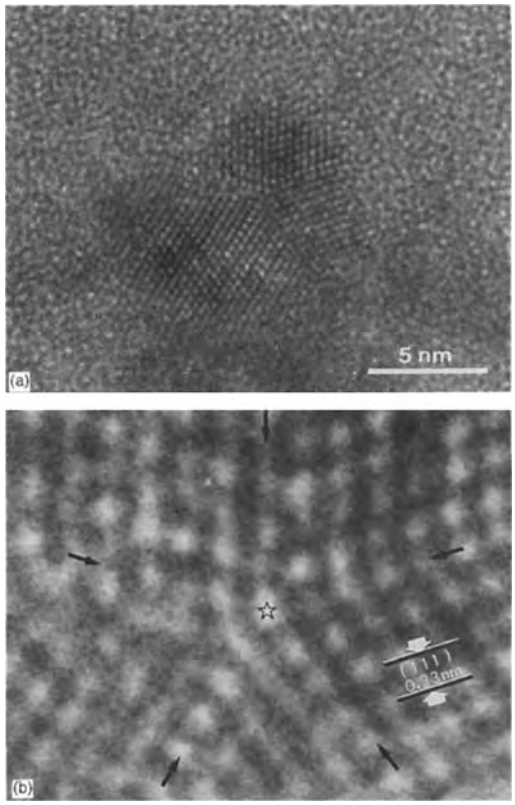


Figure 6. (a) An isolated, multiply twinned, nucleus of γ-Ge in a matrix of α-Ge; (b) the center of the nucleus showing the 0.33 nm {111} planes and the pentagonal symmetry of the {111} twin boundaries (arrows). (Courtesy of Okabe et al. [41].)

Some indication of the possibility of pentagonal, or polytetrahedral, structures in α-Ge is given by the HREM images reported by Okabe et al. [41]. What appears to be a pentagonal nucleus can be seen, from which diamond-cubic crystallites radiate (Fig. 6). Since pentagonal symmetry is inconsistent with translational periodicity, the possibility of such 'non-crystallographic' 'amorphons' has a long history [1]. However, there is relatively little experimental evidence for such structures, even though they represent a

bridge between glasses, crystals, and quasi-crystals.

Amorphous carbon is a relatively new addition to the portfolio of elemental amorphous tetrahedral semiconductors. Although 'ordinary' amorphous carbon has a structure reminiscent of crystalline graphite, it is also possible to prepare a solid with atomic characteristics similar to crystalline diamond. HREM images [42] of this amorphous tetrahedral carbon have shown distinct differences from the more graphitic solid. The fringe spacings are characteristically different and the structure is reminiscent of α-Si and Ge. (See also Chap. 1.11 of this Volume.)

A further question surrounds the lack of evidence for medium-range ordering in bright-field images of amorphous semiconductors (and oxide glasses) compared with the relative ease of observation in amorphous metals. One obvious conclusion is that amorphous metals simply are more ordered – an essential property of densely packed atoms perhaps. Another possibility is that the signal-to-noise level of HREM imaging is particularly low for low-atomic-number solids like Si and SiO_2, and that, except in special cases, the signal is almost submerged below the noise and HREM images simply reflect this.

More attention might be focused on inelastic scattering and small angle elastic scattering due to microvoids and the (rough) surfaces of a thin foil. The effect of each contribution is to add a strongly Q-dependent, small angle contribution to the noise. Graczyk and Chaudhari [43] and Patterson et al. [44] have shown that the first diffraction peaks of (relatively thick) films of α-Ge and α-Si are partly hidden by a continuous background, to which the inelastic scattering intensity I_i is the dominant contributor. Images formed from inelastically scattered electrons can be considered as an incoherent sum of beams scattered through angles determined by the energy loss [45] and are degraded compared to images formed by purely elastic scattering processes. Secondly, the low-Q incoherent inelastic scattering (involving only 'self' terms in the pair distribution function) simply contributes 'colored noise'.

The 'signal' – represented by the first peak of $S(Q)$ – is weak for α-Si and Ge compared with the amorphous metals, for which the first peak is the most prominent feature of $S(Q)$ (Fig. 1). Moreover, the 'noise' due to inelastic scattering will be weaker in α-Pd–Si alloys than in Ge or Si. Since the first peak in $S(Q)$ lies at $30\,nm^{-1}$ ($20\,nm^{-1}$ for α-Ge), and as I_i/I_c decreases rapidly with Q, inelastic scattering is relatively insignificant for amorphous metals. I_i is further reduced by the higher average atomic number of the metals: the ratio I_i/I_c decreases as Z^{-n}, where $n \simeq 1$. Taking all these effects together, Gaskell [46] has estimated that the total signal-to-noise ratio for α-Pd_4Si is likely to be larger than for α-Ge and Si by factors of at least 10 and 15, respectively.

9.1.7 Concluding Remarks

The extent of new structural information directly attributable to HREM studies is distinctly variable. There is a strong dependence on the nature of the material, on the thickness of the specimen, and, possibly, on the presence of special features, such as

a 'template' provided by the interface with a crystal that induces morphological order throughout significant thicknesses of the amorphous material. The structured images of amorphous alloys arouse the natural suspicion of surface artifacts—an inevitable consequence of the thinness of the foil. Even though α-Pd_4Si is relatively stable against crystallization and oxidation and only minor beam-induced crystallization is observed, the possibility of these effects cannot be discounted. Although it is likely that artifacts do not alter the import of the investigations cited, the possibility of oxide formation, surface metallization, or surface reconstruction has reduced the credibility of several otherwise promising measurements. Studies including perhaps in situ preparation, with careful control of the electron dose, are urgently required.

Having cited all these caveats, it is also right to stress the almost unique potential of the technique. HREM, applied to the right materials, can help to break the deadlock that still exists with regard to our interpretation of the structure of amorphous solids. There is no doubt that scattering techniques are reaching (or have reached) their limiting range (about 1 nm). There is also a growing realization that progress in the next phase will depend absolutely on an adequate understanding of the details of the medium-range structure, leading to an assessment of the relative virtues of models predicated on randomness on the one hand, and disturbed order (crystallographic or noncrystallographic) on the other. At the moment, the latter seems to be nearer the truth for most important types of glass and the former for the amorphous tetrahedral semiconductors. A truly fascinating puzzle.

9.1.8 References

[1] P. H. Gaskell in *Materials Science and Technology. A Comprehensive Review*, Vol. 9 (Ed.: J. Zarzycki), VCH, Weinheim **1991**, p. 175.

[2] J. Zarzycki, *Glasses and the Vitreous State* (Cambridge Solid State Science Series), Cambridge University Press, Cambridge **1982**.

[3] D. A. Jefferson, J. M. Thomas, G. R. Millward, K. Tsuno, A. Harriman, R. D. Brydson, *Nature* **1969**, *323*, 428.

[4] A. Howie, *J. Non-Cryst. Solids* **1978**, *31*, 41.

[5] A. Howie in *High Resolution Transmission Electron Microscopy and Associated Technique* (Eds.: P. R. Buseck, J. M. Cowley, L. Eyring), Oxford University Press, New York **1992**, p. 607.

[6] L. A. Bursill, L. G. Mallinson, S. R. Elliott, J. M. Thomas, *J. Phys. Chem.* **1981**, *85*, 3004.

[7] R. Alben, G. S. Cargill, J. Wenzel, *Phys. Rev. B* **1976**, *13*, 8358.

[8] D. J. Wallis, *Ph.D. Thesis*, University of Cambridge, Cambridge, UK **1994**.

[9] P. H. Gaskell, D. J. Wallis, *Phys. Rev. Lett.* **1996**, *76*, 66.

[10] G. Mountjoy, *Ph.D. Thesis*, University of Cambridge, Cambridge, UK **1995**.

[11] O. L. Krivanek, *Proc. Jerusalem Electron Microscopy Conference*, Tal International Publishing Company for Israeli Society for Electron Microscopy **1976**, p. 275.

[12] P. C. Shieh, C. O. Stanwood, J. M. Howe, *Ultramicroscopy* **1991**, *35*, 99.

[13] M. L. Miller, R. C. Ewing, *Ultramicroscopy* **1993**, *48*, 203.

[14] O. L. Krivanek, in *High Resolution Transmission Electron Microscopy and Associated Techniques* (Eds.: P. R. Buseck, J. M. Cowley, L. Eyring), Oxford University Press, New York **1992**, p. 519.

[15] R. Sinclair, J. Morgiel, A. S. Kirtikar, I.-W. Wu, A. Chiang, *Ultramicroscopy* **1993**, *51*, 41.

[16] G. Y. Fan, J. M. Cowley, *Ultramicroscopy* **1987**, *21*, 125.

[17] G. Y. Fan, J. M. Cowley, *Ultramicroscopy* **1988**, *22*, 49.

[18] V. Schmidt, R. Hillebrand, R. Albrecht, W. Neumann, B. Müller, *Ultramicroscopy* **1985**, *17*, 357.

[19] O. L. Krivanek, P. H. Gaskell, A. Howie, *Nature* **1976**, *262*, 454.

[20] W. O. Saxton, A. Howie, A. B. Mistry, A. Pitt in *Developments in Electron Microscopy and Analysis* (Ed.: D. L. Misell), Institute of Physics, Bristol **1977**.

[21] G. Y. Fan, J. M. Cowley, *Ultramicroscopy* **1985**, *17*, 34.

[22] R. S. Timsit, C. Gallerneault, J. C. Barry, D. Landheer, *Appl. Phys. Lett.* **1988,** *53*, 28.
[23] R. S. Timsit, W. G. Waddington, C. Gallerneault, *Ultramicroscopy* **1992,** *45*, 65.
[24] P. H. Gaskell, D. J. Smith, C. J. D. Catto, J. R. A. Cleaver, *Nature* **1979,** *281*, 465.
[25] P. H. Gaskell, D. J. Smith, *J. Microsc.* **1980,** *119*, 63.
[26] Y. Hirotsu, R. Akada, *Jpn J. Appl. Phys.* **1984,** *23*, L479.
[27] Y. Hirotsu, M. Uehara, N. Imai, *J. Electron Microsc.* **1986,** *35(Suppl.)*, 1551.
[28] K. Anazawa, Y. Hirotsu, Y. Inoue, *Acta Metallurg.* **1994,** *42*, 1997.
[29] J.-M. Dubois, P. H. Gaskell, G. Le Caer, *Proc. R. Soc. London, Ser. A* **1985,** *402*, 32329.
[30] J. L. Finney, *Proc. R. Soc. London, Ser. A* **1970,** *319*, 479.
[31] O. Beyer, C. Hoheisel, *Z. Naturforsch., Teil A* **1983,** *38*, 859.
[32] M. Shiojiri, T. Miyano, C. Kaito, *Jpn J. Appl. Phys.* **1979,** *18*, 1937.
[33] Y. Saito, C. Kaito, T. Naiki, K. Fujita, *J. Electron Microsc.* **1986,** *35(Suppl.)*, 347.
[34] P. H. Gaskell, A. B. Mistry, *Phil. Mag.* **1979** *A39*, 245.
[35] L. A. Bursill, J. M. Thomas, *J. Phys. Chem.* **1981,** *85*, 3007.
[36] D. J. Smith, W. M. Stobbs, W. O. Saxton, *Phil. Mag.* **1981,** *43*, 907.
[37] J. C. Phillips, J. C. Bean, B. A. Wilson, A. Ourmazd, *Nature* **1987,** *325*, 121.
[38] A. Ourmazd, J. C. Bean, J. C. Phillips, *Phys. Rev. Lett.* **1985,** *55*, 1599.
[39] P. H. Gaskell, A. Saeed, *J. Non-Cryst. Solids* **1988,** *106*, 250.
[40] F. Banhart, *Ultramicroscopy* **1994,** *56*, 233.
[41] T. Okabe, Y. Kagawa, S. Takai, *Phil. Mag. Lett.* **1991,** *63*, 233.
[42] A. Saeed, P. H. Gaskell, D. A. Jefferson, *Phil. Mag.* **1992,** *B66*, 171.
[43] J. F. Graczyck, P. Chaudhari, *Phys. Stat. Sol. (b)* **1973,** *58*, 163.
[44] A. M. Patterson, A. J. Craven, J. N. Chapman, A. R. Long, *Inst. Phys. Conf. Ser.* **1983,** *68*, 429.
[45] A. J. Craven, C. Colliex, *J. Microsc. Spectrosc. Electron.* **1977,** *2*, 511.
[46] P. H. Gaskell in *Electron Beam Imaging of Noncrystalline Material* (Ed.: K. M. Knowles), *IOP Short Meetings Series No. 11*, Institute of Physics, Bristol **1988,** 47.

9.2 Quasicrystalline Structures

9.2.1 Icosahedral Quasicrystals

In the search for ultrahigh-strength Al alloys, Shechtman et al. [1] tried to force more Mn into solid solution by very rapid solidification. In 1984, to their surprise they found, using selected-area electron diffraction (SAED), in an Al–14 at.% Mn alloy a metallic phase displaying icosahedral $m\bar{3}\bar{5}$ symmetry (an icosahedron is a polyhedron enclosed by 20 equilateral triangles, Fig. 1d) and no periodic translation order. The term 'quasicrystal' (QC) was coined by Levine and Steinhardt [2] for such a quasiperiodic crystal, although the term 'quasilattice' was used earlier by Mackay [3] for the two-dimensional (2-D) Penrose pattern [4] consisting of an aperiodic distribution of fat (72°) and thin (36°) rhombi and displaying five-fold symmetry (see Fig. 4b). In fact, Mackay obtained an optical transform of this pattern quite similar to the one shown in Fig. 1a and also proposed a three-dimensional (3-D) quasilattice consisting of acute ($\alpha = 63.43°$, namely, arctan2) and obtuse ($\alpha = 116.57°$) rhombohedra showing icosahedral symmetry. Icosahedral quasicrystals (IQCs) were immediately also identified by SAED in rapidly solidified Ti–Ni [5], Pd–U–Si [6], Mg–Al–Zn [7], and Al–Li–Cu [8] alloys. The era of QCs thus began. Readers interested in the fundamental characteristics of QCs are referred to *Quasicrystals: A Primer* by Janot [9] and the three volumes on *Aperiodic Crystals* edited by Jaric et al. [10].

IQC is a 3-D QC. Its characteristics, quasiperiodicity and unconventional or forbidden five-fold symmetry, are quite

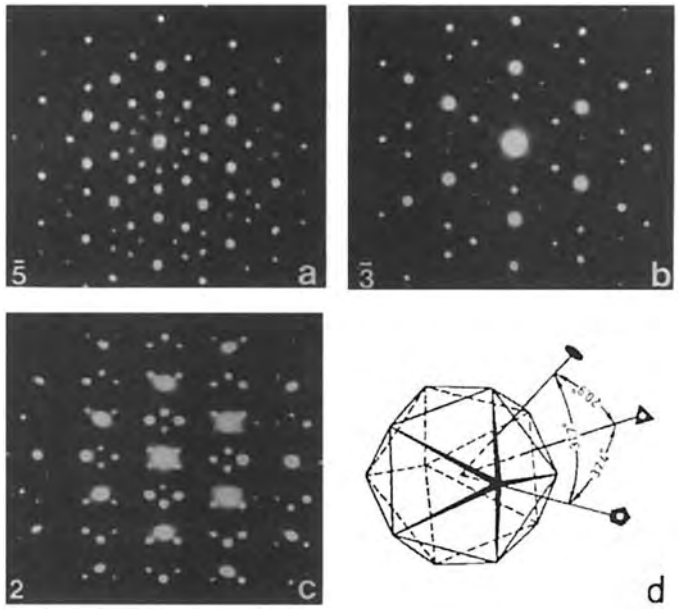


Figure 1. Electron diffraction patterns of the Ti–V–Ni icosahedral quasicrystal displaying (a) $\bar{5}$, (b) $\bar{3}$, and (c) 2-fold symmetries. (d) The $m\bar{3}\bar{5}$ point group of an icosahedron. (Reproduced from *Philos. Mag.* [5] by permission of Taylor & Francis Ltd.)

obvious in the electron diffraction patterns (EDPs) (Fig. 1). The tilted angles among the $\bar{5}$-, $\bar{3}$-, 2-fold EDPs in Fig. 1a–c agree with the angular relations of an icosahedron shown in Fig. 1d. The collinear diffraction spots within them are not equally spaced as in a crystalline substance, but show a τ distribution, such as $1{:}\tau{:}\tau^2$, where $\tau = (1+\sqrt{5})/2 = 1.61803\ldots$ is the irrational golden number associated with five-fold symmetry ($\cos(36°) = \tau/2$, $\cos(72) = (\tau-1)/2$). Moreover, the QCs formed by very rapid solidification are fine grained (of the order of μm in size) and sometimes in small amounts mixed with other phases. These factors make SAED an advantageous technique for discovering new QCs. In fact, IQCs were encountered earlier both in Al–Fe–Cu (the ψ phase) and Al–Li–Cu (the T2 phase) alloys, but their X-ray powder diffractograms, owing to the forbidden five-fold symmetry, could not be indexed then. Since the discovery of the Al–Mn IQC by SAED, a large number of IQCs have been found in binary and ternary Al alloys by means of X-ray powder diffraction methods [11].

If an IQC can be isolated from the matrix, either chemically or electrolytically, it can be examined by scanning electron microscopy (SEM). Figure 2 shows such

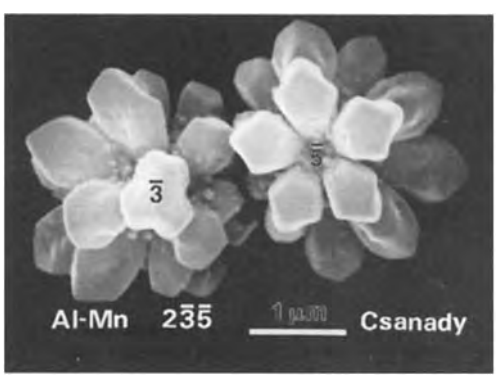


Figure 2. Scanning electron micrograph of icosahedral flowers chemically extracted from an Al–Mn alloy. (Courtesy of Dr Agnes Csanady.)

Figure 3. Scanning electron micrograph of an Al–Cu–Fe quasicrystal displaying the pentagonal dodecahedron morphology with the $m\bar{3}\bar{5}$ symmetry. (Courtesy of Dr A. P. Tsai [12] and reproduced from *Jpn J. Appl. Phys.* by permission.)

an image of the icosahedral flowers obtained by Dr Agnes Csanady from an Al–5.5 at.% Mn alloy by extraction in a solution of 2% KI in methyl alcohol. Figure 3 shows an SEM image of the pentagonal dodecahedron morphology (another regular polyhedron, enclosed by 12 regular pentagons, displaying icosahedral symmetry) of a single grain of Al–Cu–Fe IQC [12]. The $\bar{5}$- and $\bar{3}$-fold rotational symmetries, and consequently the icosahedral symmetry, are obvious in both cases. The slowly solidified Al–Li–Cu IQC shows a triacontahedron (enclosed by 30 rhombi) morphology and also exhibits icosahedral symmetry [13]. It is of interest to note that a triacontahedron contains 10 acute and 10 obtuse rhombohedra and constitutes the 3-D quasilattice [3]. Al–Cu–Fe, Al–Li–Cu, and Al–Pd–Mn IQCs are stable at high temperatures, and therefore large single crystals of millimeter or even centimeter size can be grown by slow solidification.

The five-fold symmetry of IQCs can also be shown by using high-resolution electron microscopy (HREM) imaging at the atomic scale. Figure 4a shows an HREM image of a Ti–V–Ni IQC taken along the five-fold axis [5]. Some of the

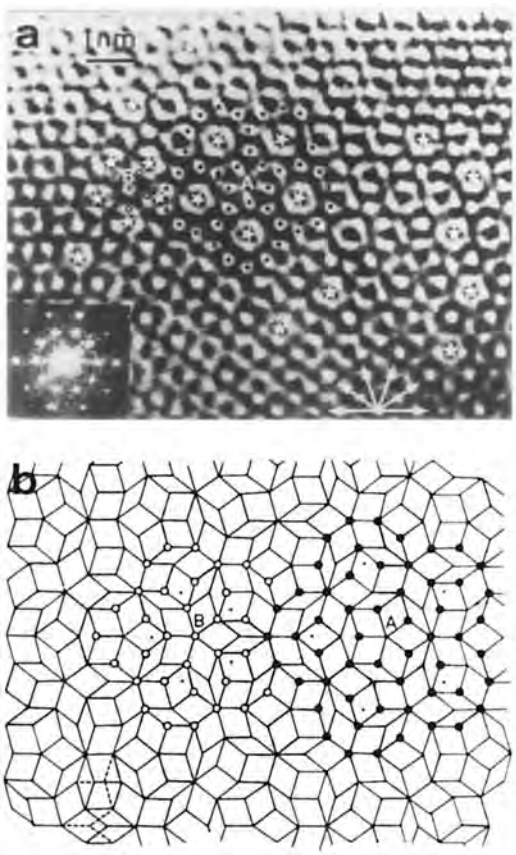


Figure 4. (a) High-resolution electron microscope image of the Ti–V–Ni icosahedral quasicrystal showing the frequent occurrence of a bull's-eye contrast surrounded by 10 bright dots (some are marked). The inset (bottom left) is the optical transform. The local five-fold configurations surrounding A and B are almost the same as those at A and B in the Penrose pattern (b). In the upper left-hand corner, an aperiodic row associated with the irrational τ is outlined; in the bottom left-hand corner the rearrangement of fat and thin rhombi in two flattened hexagons is shown by dotted lines. (Reproduced from *Philos. Mag.* [5] by permission of Taylor & Francis Ltd. (b) is adapted from A. L. Mackay, *Physica* [3].)

image points with a bull's-eye contrast are marked by stars, and the 10 bright image points surrounding some of these stars are marked by black dots. The different five-fold configurations centered on A and B agree well with the Penrose pattern given by Mackay [3] (compare A and B in Fig. 4a with the corresponding A and B in Fig. 4b). The optical transform of this HREM image shown as an inset in Fig. 4a, like the five-fold EDP given in Fig. 1a, consists of concentric decagons of spots.

9.2.2 Decagonal and Other 2-D Quasicrystals

Soon after the discovery of the Al–Mn 3-D IQC, a 2-D QC with the quasiperiodic plane perpendicular to the periodic 10-fold axis, the decagonal quasicrystal (DQC), was found coexisting with this IQC [14, 15]. If a five-fold axis of an IQC is changed to a 10-fold one, the other five-fold axes and the three-fold axes disappear, and a DQC with a periodic stacking of 2-D quasiperiodic planes results. It was shown that a block of about 0.41 nm consisting of two aperiodic layers is the basic unit, and the periodicity of DQCs can be 0.41, 0.82, 1.23, or 1.64 nm [16]. In the Al–Pd–Mn system both IQC and DQC exist, with the latter having a higher Mn and lower Pd content than the former [17]. Wollgarten et al. [18] found such a DQC lamella with a periodicity of 1.2 nm along the 10-fold axis parallel to one of the five-fold axes of the IQC matrix in an Al–Pd–Mn alloy. In order to obtain the information concerning the composition of the lamellae of a few nanometers thickness, Z-contrast imaging and the electron-energy-loss spectroscopy (EELS) techniques were used. In the Z-contrast technique the image is formed by electrons scattered by large angles, and the local brightness is proportional to the square of the mean nuclear charge Z. Such a picture of a lamella which was oriented edge on is shown in Fig. 5a. The lamella appears darker than the matrix, indicating a lower mean Z value. The EELS was recorded using a probe diameter of

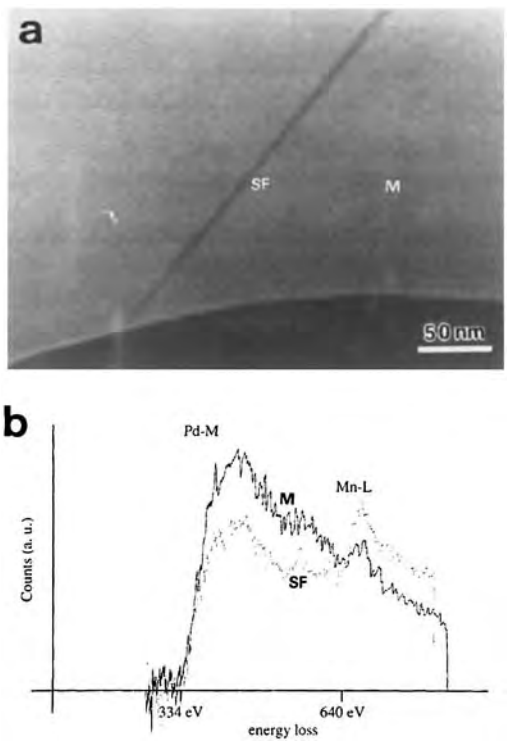


Figure 5. Z-contrast image of a lamella of a decagonal quasicrystal in a matrix of an icosahedral quasicrystal in an Al–Pd–Mn alloy. The darker contrast indicates a lower mean atomic number (a lower Pd and a higher Mn content) which is confirmed by the EELS (b). The peaks can be identified as Pd $M_{4,5}$ and the Mn $L_{2,3}$ edge. (Courtesy of Dr M. Wollgarten and reproduced from *Philos. Mag.* [18] by permission of Taylor & Francis Ltd.)

0.3 nm in a scanning transmission electron microscope (STEM). The solid and broken curves in Fig. 5b show the EELS spectra recorded from the matrix and the lamella, respectively. Evaluating the peaks of the Pd $M_{4,5}$ and the Mn $L_{2,3}$ edge by standard methods yields the result that the palladium content was about 67% (±13%) and the Mn content about 200% (±50%) of that of the icosahedral matrix. The stoichiometry of the matrix was determined by energy-dispersive X-ray microscopy (EDX) measurements to be $Al_{74}Pd_{16}Mn_{10}$. The EELS results therefore indicate an approximate composition of the lamella of $Al_{69}Pd_{11}Mn_{20}$. This is close to the composition of the Al–Pd–Mn alloy in which DQC has been found [17].

A stable DQC [19] with a decaprism morphology growing along the 10-fold axis [20, 21] (see the SEM image in Fig. 6) has been found in slowly solidified Al–Cu–Co alloys. Its 10-fold ED pattern shows a 10-fold aperiodic distribution of diffraction spots. The bright image points in the HREM image show a periodic arrangement along the 10-fold axis and an aperiodic arrangement with a τ relationship along the two-fold axis (both marked on the HREM image shown in Fig. 6).

Owing to the two-fold symmetry inherited in an SAED pattern, the five-fold EDP of a DQC with the point group $\bar{1}0m2$, where $\bar{1}0$ is equivalent to 5/m, shows a 10-fold distribution of electron diffraction spots. However, this is not the case with the convergent-beam electron diffraction (CBED) pattern which shows the real symmetry of a substance in the direction of the electron beam. In the CBED pattern of the Al–Cu–Co DQC shown in Fig. 7,

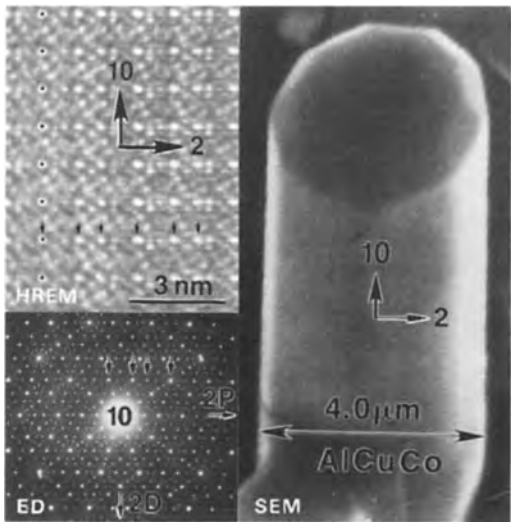


Figure 6. SEM image showing the decaprism growth morphology along the 10-fold axis of the Al–Cu–Co decagonal quasicrystal. The 10-fold electron diffraction (ED) pattern and the HREM image taken along a two-fold axis showing the periodic and aperiodic arrangements along the 10-fold 10 and two-fold 2P directions, respectively, are also shown. (Courtesy of Dr L. X. He.)

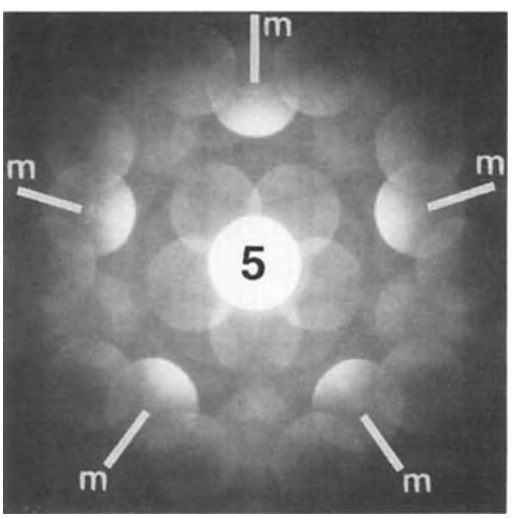


Figure 7. Convergent-beam electron diffraction pattern of an Al–Cu–Co decagonal quasicrystal showing its 5m symmetry. (Courtesy of Dr M. Tanaka and reproduced from *Mater. Sci. Eng.* [22] by permission of Elsevier.)

the discs clearly show 5m symmetry with mirror planes parallel to the 5 axis (space group $\bar{1}0m2 = 5/mm2$) in the Al DQC [22]. However, DQCs with the point group 10mm (space group $P10_5/mmc$) have also been found by CBED in several alloy systems, such as $Al_{70}Ni_{20}Fe_{10}$ [23]. Moreover, CBED is also very effective for showing any departure from the true five-fold or ten-fold symmetry when there are a large amount of defects in a QC [24].

The Fourier-filtered scanning tunneling microscopy (STM) image of the Al–Cu–Co DQC decaprism with a cross-section of a few millimeters is shown in Fig. 8 [25]. It looks exactly the same as the HREM image taken along the 10-fold axis of this DQC [26]. Both images show an aperiodic arrangement of 'decagon wheels' consisting of 10 bright image points surrounding a bull's-eye contrast at the center. The STM image shows the true surface structure, whereas the HREM image results from the projection of a 3-D structure along the electron beam. Some objections have been raised to the use of the projected HREM image to explain the 3-D structure. The bright image points in the STM image were shown to represent Cu/Co atoms at the center of pentagons in the Penrose pattern [25]. This proves the quasiperiodic nature of QCs and thus rules out the possibility of five-, ten-, or icosa-fold twins advocated by Pauling [27]. Moreover, this result also shows that one may use the HREM image to interpret the bulk quasicrystalline structure, especially if the HREM images compare well with simulated ones. Figure 8 also shows several quasiperiodic layers with an atomic step of about 0.21 nm between any two of them.

Figure 8. Scanning tunneling microscope image of the surface structure of an Al–Cu–Co decagonal crystal. Ten-fold and quasiperiodic arrangements on several atomic layers are obvious. (Courtesy of Drs A. R. Kortan and H. S. Chen and reproduced from *Phys. Rev. Lett.* [25].)

In addition to the DQCs with a 10-fold rotational axis, other types of 2-D QCs, such as the octagonal QCs with an eight-fold rotational axis [28] and dodecagonal QCs with a 12-fold rotational axis [29], have also been found using SAED and HREM. The octagonal quasilattice was shown by HREM to consist of squares and 45° rhombi associated with the irrational $\sqrt{2}$, whereas the dodecagonal quasilattice of squares 60° and 30° rhombi was associated with $\sqrt{3}$. Furthermore, one-dimensional (1-D) QCs with periodic layers stacked aperiodically have also been found [30].

9.2.3 Quasicrystalline Structures

From the above discussion it is clear that quasiperiodicity is a fundamental property of a QC. For instance, if the length of the side of the fat or thin rhombi is considered to be unity, then the spacings between two lattice points along the oblique direction at the upper-left-hand corner in Fig. 4b are $\tau:1:1:\tau:\tau^{-1}:1:\tau:1$. This is an aperiodic series of 1 and the irrational τ, and one needs two lengths to describe the positions of the 1-D lattice points. In a 3-D QC, there will be three such series, and one needs six basic vectors to describe the position of a quasilattice point [31–33]. Hence, IQCs can be considered as periodic in the six-dimensional space and the formula used to calculate the structure factors, diffraction intensity, image contrast, HREM simulation etc., for crystals can principally be applied to a QC if the position vector $\boldsymbol{r}$ in the real space and the reciprocal vector $\boldsymbol{g}$ in the reciprocal space are raised from the 3-D to the 6D space. Of course, the calculated results in the 6D space have to be projected back to the 3-D physical space. As in incommensurate crystals, this mathematical treatment will materially simplify the study of quasicrystalline structures. In fact, QCs are a special kind of incommensurate structure, with unconventional rotational symmetries and no average lattice.

Five-fold symmetry and any other rotational symmetry above six-fold are forbidden in a periodic lattice, but not in a quasiperiodic lattice, such as the 2-D Penrose lattice. In order to ensure an aperiodic tiling, there are strict local rules governing the tiling of the fat and thin rhombi in Fig. 4b or the 3-D acute and obtuse rhombohedra. In other words, there is strict position order in a quasilattice, although it is aperiodic. Therefore, high-quality QCs can give as sharp X-ray reflections or electron diffraction spots as can crystals.

If the 6D superspace is divided into two orthogonal 3-D spaces, one parallel or physical space $E^{\parallel}$ and one perpendicular or pseudo space $E^{\perp}$. For simplicity, this can be illustrated in the 2-D–1-D case. Figure 9a shows the irrational projection of a 2-D square lattice E onto the 1-D parallel space $E^{\parallel}$ [34]. If the lattice points within the stripe with an irrational slope of $\tan\alpha = \tau$ and a width of the projected diagonal of the square on $E^{\perp}$ are chosen, the projected line segments will follow the $LSLLSLSL\ldots$, where $L = \sin\alpha$, $S = \cos\alpha$, and $L/S = \tan\alpha = \tau$. This is the famous Fibonacci series associated with the golden number τ. This irrational number can be successively approximated by a ratio of F_{n+1}/F_n, where $F_{n+1} = F_n + F_{n-1}$, $F_0 = 0$, $F_1 = 1$, such as 1/0, 2/1, 3/2, 5/3, 8/5, 13/8,.... If the stripe in Fig. 9a is

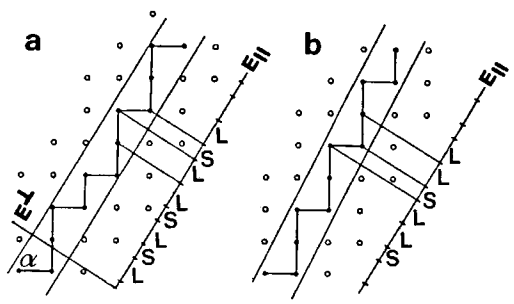


Figure 9. Projecting a stripe of a 2-D square lattice (a) irrationally ($\tan\alpha = \tau$, the golden number) onto a 1-D $E^{\parallel}$ space to yield an aperiodic Fibonacci series, or (b) rationally ($\tan\alpha = 2$) to yield a periodic lattice with a repeat unit of *LLS*. The latter results from a 2/1 approximation of the irrational τ. (Adapted from Elser and Henley in *Phys. Rev. Lett.* [34].)

rotated slightly and the slope becomes rational, say 2/1, then the projected line segments will be a 1-D periodic lattice, namely, *LLSLLS*... (Fig. 9b). Obviously, both 1-D periodic and aperiodic lattices can be embedded in a 2-D periodic lattice, depending on whether the slope of the projection stripe is rational or irrational. From this point of view, a QC can be considered as a crystal with an infinitely large unit cell, since the ratio of F_{n+1}/F_n approaches τ when n approaches infinity. Or, in other words, a QC is the end member of a series of crystalline phases with τ inflations. Recently, a series of monoclinic $Al_{13}Co_4$ phases with their lattice parameters a and c inflated to about τ, τ^2, τ^3, and τ^4 have been found [35]. These inflated phases have similar pentagonal layers, as in $Al_{13}Co_4$ and the Al–Co DQC. Such crystalline phases are called 'crystalline approximants' or simply 'approximants' [34]. The local structure and composition in a QC and its approximant are alike. However, the subunits are arranged periodically in an approximant and aperiodically in a QC.

Thus, the five-fold EDP and the simulated HREM image of the Al–Mn IQC can be calculated using a 6-D approach which is then compared with the experimental results. When indexing of the diffraction spots or reflections, we need consider only the 3-D quasilattice or the 6-D periodic lattice. There are several 6-D indexing systems [33, 36, 37]. However, in order to calculate the diffraction intensities and the image contrast, the 6-D periodic unit cell has to be decorated with 6-D superatoms. The 6-D structural model of a QC can be constructed from the structure of its 3-D approximant, such as body-centered cubic (b.c.c.) α-(AlMnSi) [34] and b.c.c. $Mg_{32}(Al,Zn)_{49}$ [38]. The simulated HREM images of the Al–Mn IQC have been calculated both from the bare quasilattice and from the quasiperiodic structural model using the Mackay double icosahedra found in cubic α-(AlMnSi) [39].

Simulated HREM images have also been calculated using the 3-D multislice method, either from a huge quasicrystalline structural model consisting of 59 567 atoms [40], or from a large unit cell (3.88 nm × 3.30 nm) by using a periodic extension method [41].

9.2.4 Defects in Quasicrystals

As there are two or three kinds of unit cell in a QC, one has to consider the tiling mistakes which do not exist in crystals. For instance, the fat and thin rhombi in the two flattened hexagons at the bottom-left-hand corner in Fig. 4b can also be tiled differently, as shown by the dotted lines. This does not change the shape of the hexagons and thus causes no distortion [42]. However, the position of some lattice points in this 2-D quasilattice is changed, and consequently the phase in the density-wave description of the QC will change accordingly. Following the terminology used for incommensurate crystals, this kind of defect is called a 'phason strain' or simply a 'phason' [43, 44]; the displacement strain is called a 'phonon'.

Both whole and partial dislocations have been observed in QCs, and diffraction contrast experiments using the criterion $\boldsymbol{g}^{\parallel} \cdot \boldsymbol{b}^{\parallel} = 0$ have been carried out in order to determine their Burgers vector $\boldsymbol{b}$ in the

physical space. However, complications arise because $\boldsymbol{g}^{\parallel} \cdot \boldsymbol{b}^{\parallel} = 0$ does not mean $\boldsymbol{g} \cdot \boldsymbol{b} = 0$. Both $\boldsymbol{g}$ and $\boldsymbol{b}$ are now in the 6-D space and the extinction criterion $\boldsymbol{g} \cdot \boldsymbol{b} = 0$ contains two terms

$$\boldsymbol{g} \cdot \boldsymbol{b} = \boldsymbol{g}^{\parallel} \cdot \boldsymbol{b}^{\parallel} + \boldsymbol{g}^{\perp} \cdot \boldsymbol{b}^{\perp} = 0$$

Woolgarten et al. [45] differentiated two cases:

Strong extinction condition:

$$\boldsymbol{g}^{\parallel} \cdot \boldsymbol{b}^{\parallel} = \boldsymbol{g}^{\perp} \cdot \boldsymbol{b}^{\perp} = 0$$

Weak extinction condition:

$$\boldsymbol{g}^{\parallel} \cdot \boldsymbol{b}^{\parallel} = -\boldsymbol{g}^{\perp} \cdot \boldsymbol{b}^{\perp} \neq 0$$

The disappearance of the diffraction contrast of a dislocation under both extinction conditions has been verified experimentally [18, 46]. This shows that a dislocation in a QC causes both phonon and phason strains. However, the extinction criterion in the diffraction-contrast technique gives only the direction of the Burgers vector of a dislocation and provides no information concerning its magnitude. Moreover, sometimes it is not easy to find suitable reflections to satisfying the above extinction conditions.

Wang and coworkers [47–49] extended the defocus CBED technique of Cherns and Preston [50] to determine both the direction and magnitude of the Burgers vector $\boldsymbol{b}$ of a dislocation in a QC if $\boldsymbol{g} \cdot \boldsymbol{b} = n$ is treated in 6-D space. In the defocus CBED pattern, a high-order reflection fringe in the $\boldsymbol{g}$ diffracted disc and its corresponding deficient line in the transmitted disc, when crossing a dislocation with a Burgers vector $\boldsymbol{b}$, will split into $n + 1$ sublines with n nodes. The sense of $\boldsymbol{g} \cdot \boldsymbol{b}$ will be positive (or negative), if on the $\boldsymbol{u} \times \boldsymbol{c}$ side of the dislocation line of the reflection fringe twists along (or opposite to) the $\boldsymbol{g}^{\parallel}$ direction, where $\boldsymbol{u}$ is the line direction of the dislocation and $\boldsymbol{c}$ is a vector pointing to the incident beam cross-over from the dislocation line. When the splitting behavior of six reflection fringes with linearly independent indices (in the simplest case they may be chosen from three linearly independent reflection rows) is observed using the defocus CBED technique, six independent equations $\boldsymbol{g}_i \cdot \boldsymbol{b} = n_i$ $(i = 1, 2, \ldots, 6)$ will result, and from these the 6-D Burgers vector $\boldsymbol{b}$ can be obtained.

Figure 10 shows a defocus CBED pattern taken from a dislocation line (denoted by $\boldsymbol{u}$) in an Al–Cu–Fe IQC [49]. Six high-order Laue zone (HOLZ) lines (marked 1–6) crossed by the dislocation line are split into 3, 2, 3, 2, 3, and 2 sublines, respectively. The corresponding values of $\boldsymbol{g} \cdot \boldsymbol{b}$ are determined to be -2, -1, 2, -1, -2, and 1, respectively. After indexing these six $\boldsymbol{g}$ values, let $\boldsymbol{b} = [b_1, b_2, b_3, b_4, b_5, b_6]$. A set of six linear

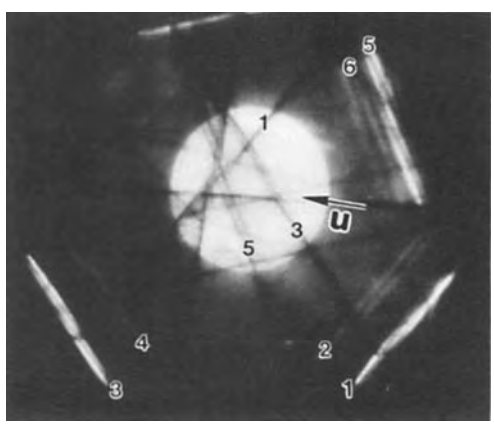


Figure 10. Defocus CBED pattern showing the split of the bright diffracted and the dark transmitted fringes into sublines caused by a dislocation line (direction $\boldsymbol{u}$) in an Al–Cu–Fe icosahedral quasicrystal. (Courtesy of Dr R. Wang and reproduced from *Phil. Mag.* [49] by permission of Taylor & Francis Ltd.)

equations with six parameters is obtained:

$$\begin{bmatrix} -4 & -6 & -8 & -2 & 4 & 2 \\ -3 & -3 & -5 & -1 & 3 & 1 \\ -2 & -4 & 4 & 6 & -2 & -8 \\ -1 & -3 & 3 & 3 & -1 & -5 \\ 0 & 2 & -4 & -4 & 2 & 6 \\ 1 & 1 & -3 & -3 & 1 & 3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix} = \begin{bmatrix} -2 \\ -1 \\ 2 \\ -1 \\ -2 \\ 1 \end{bmatrix}$$

The Burgers vector $\boldsymbol{b}$ is found to be $\boldsymbol{b}^5 = [1, 1, -2, 1, -1, -1]$ in the direction of a five-fold axis of the IQC along an edge of the acute rhombohedron. The Burgers vector components in the physical and pseudo spaces are irrational: $\boldsymbol{b}^{5\parallel} = (3 - 2\tau, 2 - \tau, 0)$ and $\boldsymbol{b}^{5\perp} = (-2 - 3\tau, 0, -1 - 2\tau)$. Similarly, dislocations of the three-fold type with Burgers vectors $\boldsymbol{b}^3$ in the direction of a three-fold axis along the body diagonal of the acute rhombohedron [49] and of the two-fold type with $\boldsymbol{b}^2$ in the direction of a two-fold axis along the face diagonal of the acute rhombohedron [47, 48] have been found.

Phason strains will modify or change the phase term of the wave-density expression of a QC. In real space this corresponds to tiling mistakes or a rearrangement of the tiles [42], such as the transformation of *LSLLSLS*... into *LLSLLSLL*... in Fig. 9. This is in fact the transformation of a Fibonacci aperiodic association with τ into a 2/1 approximant series. In reciprocal space, phason strains will cause a movement of the diffraction spots of a QC towards the positions corresponding to the approximant [34]. Phasons and the corresponding transformation of a QC to an approximant has been studied using either electron diffraction or HREM in octagonal QCs [51–53], DQCs [26, 54–58] and IQCs [59–62]. Moreover, domains and discommensurations have also been observed in octagonal QCs [63, 64] and in DQCs [22, 23].

Readers interested in the electron diffraction and microscopy of QCs are referred to some recent reviews [65, 66] and edited volumes [67–70].

Acknowledgements

Thanks are due to Drs H. S. Chen, Agnes Csanady, L. X. He, A. R. Kortan, M. Tanaka, A. P. Tsai, R. Wang, and M. Wollgarten for supplying original prints for the figures.

9.2.5 References

[1] D. Shechtman, I. Blech, D. Gratias, J. W. Cahn, *Phys. Rev. Lett.* **1984,** *53*, 1951.
[2] D. Levine, P. J. Steinhardt, *Phys. Rev. Lett.* **1984,** *53*, 2477.
[3] A. Mackay, *Physica* **1982,** *114A*, 609.
[4] R. Penrose, *Bull. Inst. Math. Appl.* **1974,** *10*, 266.
[5] Z. Zhang, H. Q. Ye, K. H. Kuo, *Philos. Mag.* **1985,** *52*, L49.
[6] S. J. Poon, A. Drehman, K. R. Lawless, *Phys. Rev. Lett.* **1985,** *55*, 2324.
[7] P. Ramanchandrarao, G. V. S. Sastry, *Pramana* **1985,** *25*, L225.
[8] P. Sainfort, B. Dubost, A. Dubus, *C.R. Acad. Sci. Paris, Ser. II* **1985,** *301*, 689.
[9] C. Janot, *Quasicrystals: A Primer*, Oxford University Press, Oxford **1992.**

[10] M. V. Jaric (Ed.), *Aperiodicity and Order*: Vol. 1, *Introduction to Quasicrystals*, Academy Press, Boston **1988**; Vol. 2, *Introduction to the Mathematics of Quasicrystals*, Academy Press, Boston **1989**; M. V. Jaric and D. Gratias (Eds.), Vol. 3, *Extended Icosahedral Structures*, Academy Press, Boston **1989**.

[11] P. A. Bancel, P. A. Heiney, *Phys. Rev. B* **1986,** *33*, 7917.

[12] A.-P. Tsai, A. Inoue, T. Masumoto, *Jpn J. Appl. Phys.* **1987,** *26*, L1505.

[13] F. W. Gayle, *J. Mater. Res.* **1987,** *2*, 1.

[14] L. Bendersky, *Phys. Rev. Lett.* **1985,** *55*, 1461.

[15] K. Chattopadhyay, S. Ranganathan, S. Subbanna, N. Thangaraj, *Script. Metall.* **1985,** *19*, 767.

[16] X. Z. Li, K. H. Kuo, *Philos. Mag. Lett.* **1988,** *58*, 167.

[17] C. Beeli, H. U. Nissen, J. Rabadey, *Philos. Mag. Lett.* **1991,** *63*, 87.

[18] M. Wollgarten, H. Lakner, K. Urban, *Philos. Mag. Lett.* **1993,** *67*, 9.

[19] L. X. He, Z. Zhang, Y. K. Wu, K. H. Kuo, *Int. Phys. Conf. Ser. 93* **1988,** *2*, 501.

[20] L. X. He, Y. K. Wu, X. M. Meng, K. H. Kuo, *Philos. Mag. Lett.* **1990,** *61*, 15.

[21] A. R. Kortan, F. A. Thiel, S. H. Chen, A. P. Tsai, A. Inoue, T. Masumoto, *Phys. Rev. B* **1989,** *40*, 9398.

[22] K. Saitoh, K. Tsuda, M. Tanaka, A. P. Tsai, A. Inoue, T. Masumoto, *Mater. Sci. Eng.* **1994,** *A181/A182*, 805.

[23] M. Tanaka, K. Tsuda, M. Terauchi, A. Fujiwara, A. P. Tsai, A. Inoue, T. Masumoto, *J. Non-Crystalline Solids* **1993,** *153/154*, 98.

[24] M. Tanaka, M. Terauchi, K. Hiraga, M. Hirabayashi, *Ultramicroscopy* **1985,** *17*, 279.

[25] A. R. Kortan, R. S. Becker, F. A. Thiel, H. S. Chen, *Phys. Rev. Lett.* **1990,** *64*, 200.

[26] H. L. Li, Z. Zhang, K. H. Kuo, *Phys. Rev. B* **1994,** *60*, 3645.

[27] L. Pauling, *Nature* **1985,** *317*, 512.

[28] N. Wang, H. Chen, K. H. Kuo, *Phys. Rev. Lett.* **1987,** *59*, 1010.

[29] T. Ishimasa, H. U. Nissen, Y. Fukano, *Phys. Rev. Lett.* **1985,** *55*, 511.

[30] L. X. He, X. Z. Li, Z. Zhang, K. H. Kuo, *Phys. Rev. Lett.* **1988,** *61*, 1116.

[31] P. A. Kalugin, A. Yu. Kitayev, L. S. Levitov, *J. Phys. Lett.* **1885,** *46*, L601.

[32] A. Katz, M. Duneau, *J. Phys.* **1986,** *47*, 181.

[33] V. Elser, *Phys. Rev. B* **1985,** *32*, 4892.

[34] V. Elser, C. L. Henley, *Phys. Rev. Lett.* **1985,** *55*, 2883.

[35] X. L. Ma, X. Z. Li, K. H. Kuo, *Acta Crystallogr., Sect. B* **1995,** *51*, 36.

[36] P. A. Bancel, P. A. Heiney, P. W. Stephens, A. I. Goldman, P. M. Horn, *Phys. Rev. Lett.* **1985,** *54*, 2422.

[37] J. W. Cahn, D. Shechtman, D. Gratias, *J. Mater. Res.* **1986,** *1*, 13.

[38] C. L. Henley, V. Elser, *Phil. Mag. B* **1986,** *53*, L59.

[39] M. Cornier, K. Zhang, R. Portier, D. Gratias, *J. Phys.* **1987,** *47*, colloque C3-447.

[40] W. Krakow, D. P. DiVincenzo, P. A. Bancel, E. Cockayne, V. Elser, *J. Mater. Res.* **1993,** *8*, 24.

[41] C. Beeli, S. Horiuchi, *Philos. Mag. B* **1994,** *70*, 215.

[42] V. Elser, *Phys. Rev. Lett.* **1985,** *54*, 1730.

[43] P. Bak, *Phys. Rev. Lett.* **1985,** *54*, 1517.

[44] D. Levine, T. C. Lubensky, S. Ostlund, S. Ramaswamy, P. J. Steinhardt, J. Toner, *Phys. Rev. Lett.* **1985,** *54*, 1520.

[45] M. Wollgarten, D. Gratias, Z. Zhang, K. Urban, *Phil. Mag. A* **1991,** *64*, 819.

[46] M. Wollgarten, Z. Zhang, K. Urban, *Phil. Mag. Lett.* **1992,** *65*, 1.

[47] R. Wang, M. X. Dai, *Phys. Rev. B* **1993,** *47*, 15326.

[48] J. L. Feng, R. Wang, *Phil. Mag. A* **1994,** *69*, 981.

[49] J. L. Feng, R. Wang, *Phil. Mag. Lett.* **1994,** *69*, 309.

[50] D. Cherns, A. R. Preston in *Proc. XI Int. Congress on Electron Microscopy* (Eds.: T. Imura, S. Maruse, T. Suzuki), Japan Society Electron Microscopy, Tokyo **1986,** pp. 721–722.

[51] Z. M. Wang, K. H. Kuo, *Acta Crystallogr., Sect. A* **1988,** *44*, 857.

[52] N. Wang, H. Kuo, *Phil. Mag. Lett.* **1990,** *61*, 63.

[53] J. C. Jiang, K. H. Kuo, *Ultramicroscopy* **1994,** *54*, 215.

[54] K. Hiraga, *J. Non-Crystalline Solids* **1993,** *153/154*, 28.

[55] H. Zhang, K. H. Kuo, *Phys. Rev.* **1990,** *42*, 8907.

[56] L. E. Levine, P. C. Gibbons, A. M. Viano, *Phil. Mag. B* **1994,** *70*, 11.

[57] K. Edagawa, M. Ichihara, K. Suzuki, S. Takeuchi, *J. Non-Crystalline Solids* **1993,** *153/154*, 19.

[58] G. Van Tendeloo, J. Van Landuyt, S. Amelinckx, S. Ranganathan, *J. Microsc.* **1988,** *149*, 1.

[59] K. Hiraga, *J. Electron Microsc. (Jpn)* **1991,** *40*, 81.

[60] F. H. Li, G. Z. Pan, S. Z. Tao, M. J. Hui, Z. H. Mai, X. S. Chen, L. Y. Cai, *Phil. Mag. B* **1990,** *29*, 535.

[61] F. H. Li, G. Z. Pan, D. X. Huang, H. Hashimoto, Y. Yokota, *Ultramicroscopy* **1992,** *45*, 299.

[62] X. Zhang, R. M. Stroud, J. L. Lippert, K. F. Kelton, *Phil. Mag. B* **1994,** *70*, 927.

[63] J. C. Jiang, N. Wang, K. K. Fung, K. H. Kuo, *Phys. Rev. Lett.* **191,** *67*, 1302.

[64] J. C. Jiang, K. K. Fung, K. H. Kuo, *Phys. Rev. Lett.* **1992,** *68*, 616.
[65] K. H. Kuo, *Mater. Chem. Phys.* **1994,** *39*, 1.
[66] K. F. Kelton, in *Intermetallic Compounds* (Eds.: J. H. Westbrook, R. L. Fleischer), Wiley, New York **1994,** Chap. 20.
[67] M. J. Yacaman, M. Torres (Eds.), *Crystal–Quasicrystal Transition*, North-Holland, Amsterdam **1993**.
[68] K. F. Kelton, P. C. Gibbons, A. I. Goldman (Eds.), *J. Non-Crystalline Solids* **1993,** *153/154*, 1.
[69] K. H. Kuo, S. Takeuchi (Eds.), *Mater. Sci. Forum* **1994,** *151/152*, 1.
[70] T. Masumoto, K. Hashimoto (Eds.), *Mater. Sci. Eng.* **1994,** *A181/A182*, 717.

Handbook of Microscopy
S. Amelinckx, D. van Dyck,J. van Landuyt , G. van Tendeloo

10 Medical and Dental Materials

10.1 Introduction

Generally, the term of biomaterials means materials that are used for restoration of functions which were destroyed by a disease or by an accident in the human body. For instance, it is well known that gold is one of the oldest biomaterials used in dentistry. Figure 1 exhibits the early Phoenician dental restoration which was to maintain teeth in a fixed position in the mouth known as the Gaillardot specimen, dating from about 600–400 BC. Ambrose Paré, who has often been called 'the father of surgery', described a palatal obturator which was a device to fill a perforation of the hard palate in his *Ten Books of Surgery* (1563). This device consisted of a curved sheet of gold large enough in size to cover the hole and shaped to conform to the configuration of the mouth.

In the 1770s, surgeons were in dispute over the manner by which to fix a fractured bone. Iron wire was used although it was very difficult to evaluate whether the use of metal increased the incidence of infection. The first study of biocompatibility of metals to tissue was performed using platinum, gold, silver, lead, etc. in a series of experiments with dogs in 1829 undertaken by Levert [1]. Various metals and alloys were tested in the nineteenth century and the reaction of metals to tissue was recognized in the early 1900s. To provide a successful application of an implant, it was clear that the implant must be put into the body without bacterial contamination but also that an adequate combination of metal components should be selected to avoid the disintegration due to electrolytic reaction. Apart from this basic understanding, the knowledge of biomaterials developed on a basis of trial and error over many decades. Today, biomaterials science is a recognized interdisciplinary field bringing together medical, dental and materials science.

Table 1 lists conventional biomaterials, used currently, which are classified by their functions and applications. Such biomaterials are roughly divided into two categories: structural and functional materials. Structural biomaterials, on the one hand, are used for withstanding cyclic loading, such as mastication or walking motions. Functional biomaterials, on the other hand, are applied to realize a specific biofunctional purpose, because of their peculiar characteristics, such as a drug delivery system with the degradation or bioerodability, a hemidialysis membrane with permeability. Furthermore, these biomaterials may be implanted either

Figure 1. Retentive prosthesis supposedly of Phoenician origin dating from 600–400 BC. (Photo by K. Yasuda with permission from the Louvre Museum, Paris.)

entirely within the soft and/or hard tissues (e.g., a bone plate, a hip joint) or partially in the tissue but with penetration of an epithelial surface (e.g., intraoral mucosae, such as dental implants, because a part of the device is exposed to the exterior of the body). They may also be placed on an epithelial surface (e.g., a denture).

Essentially, the requirements of biomaterials are that there should be a lack of significant interaction between the material and the surrounding biological system of the human body. That is, they should be nontoxic, nonirritant, noncarcinogenic, and nonallergenic as well as providing suitable mechanical properties such as strength, rigidity, elasticity, and high chemical stability in an aggressive environment inside of the human body. These requirements are considered to be divided into two parts at an interface between the material and surrounding biological system (e.g., tissue, blood, extracellular fluids, etc.) because a clinical application of the materials involves the creation of the interface. Figure 2 shows a schematic illustration of structural aspects depending on the size of biomaterial and biological system at the interface [2]. For example, surface roughness of an implant device may affect cell behavior at the 10 μm scale, since adsorptive and migratory behavior of cells play an important role in forming a functional seal with percutaneous devices which must penetrate through a stratified epithelium, such as dental implants, auditory prostheses, catheters for dialysis, etc. An oxide layer of passivating metals and alloys also plays an important role in resisting corrosion in physiological environments.

Because of the preference for energetically favorable reaction sites, which are the emergence sites of dislocations on a surface or the various interfaces intersected at a surface and along grain boundaries, corrosion is induced preferentially at these surface structures. When a metallic implant meets with tissue, an interaction takes place at the interface between the surface atoms of the oxide layer and the molecular components of the biological system. Consequently, cations are released from the oxide layer into the body fluids, and are able to bind to proteins in the biological system which will induce an effect on tissue. Thus, the study of biomaterials science is carried out by considering the interface from two quite different aspects. First, biomaterials science itself deals with matters such as the specific suitability of mechanical properties, corrosion resistance and other characteristics necessary to realize functional performance within the body. Second, the effects of the chosen biomaterials on the biological system in the body forms the study of biocompatibility.

Table 1. Classification of medical and dental materials according to category, application and function.

Biomaterials	Principal applications	Function applied
Metals and alloys		
316 stainless steel	Joint replacement	Strength and wear resistance
	Spinal instrumentation	Toughness
	Fracture fixation	Rigidity
Titanium,	Joint replacement; tibial components	Strength and wear resistance
Ti–6Al–4V,	Fracture fixation;	Rigidity
Ti–6Al–7N	Pacemaker encapsulation	Strength and biocompatibility
Co–Cr alloys	Joint replacement;	Strength and wear resistance
	Dental implants,	Toughness
	Dental prostheses	Strength
	Orthodontics	Resilience
	Heart valves	Toughness
Gold alloys	Dental restorations; fillings	Ductility
	Dental restorations; prostheses	Strength
	Dental restoration; bridges	Resilience
Platinum and PGM	Electrodes	Conductivity and corrosion resistance
Hg–Ag–Sn amalgam	Dental restorations	Plasticity and hardness
Ceramics and glasses		
Alumina	Joint replacement; femoral stems	Strength and inertness
	Acetabular cups	Wear resistance
	Dental implants	Strength and rigidity
Calcium phosphates	Ceramic femoral stems	Strength and inertness
	Scaffold material bone repair	Biological affinity
	Surface coating on metals	Biological affinity
Bioactive glasses	Bone replacement	Biological affinity
Porcelain	Dental restorations	Hardness and inertness
Polymers		
Polyethylene (UHMWPE)	Joint replacement	Wear resistance
Polytetrafluoroethylene (PTFE)	Soft-tissue augmentation	Biocompatibility
	Membrane for blood-gas exchangers	Permeability and biocompatibility
	Vascular prostheses	Antithrombotic
Polypropylene	Membrane oxygenator	Permeability and biocompatibility
Polyesters	Vascular prostheses	Antithrombotic
	Drug delivery systems	Bioerodibility
	Absorbable sutures	Bioerodability
Polyurethanes	Blood-contacting devices	Antithrombotic
Polyvinylchloride (PVC)	Tubing	Antithrombotic
Polymethylmethacrylate (PMMA)	Dentures	Strength and biocompatibility
	Intraocular lenses	Flexibility and gas permeability
	Joint replacement	Strength
Silicones	Soft-tissue replacement	
Hydrogels	Ophthalmology	Optical characteristics
	Blood-contacting devices	Hemo-compatibility
Composites		
Dimethylacrylate-quartz	Dental restorations	Compressive strength
Carbon fiber – thermosetting resins	Bone repair	Toughness and biocompatibility
Carbon fiber – thermoplastics	Bone repair	Toughness and biocompatibility
Carbon–carbon	Bone and joint replacement	Strength and toughness

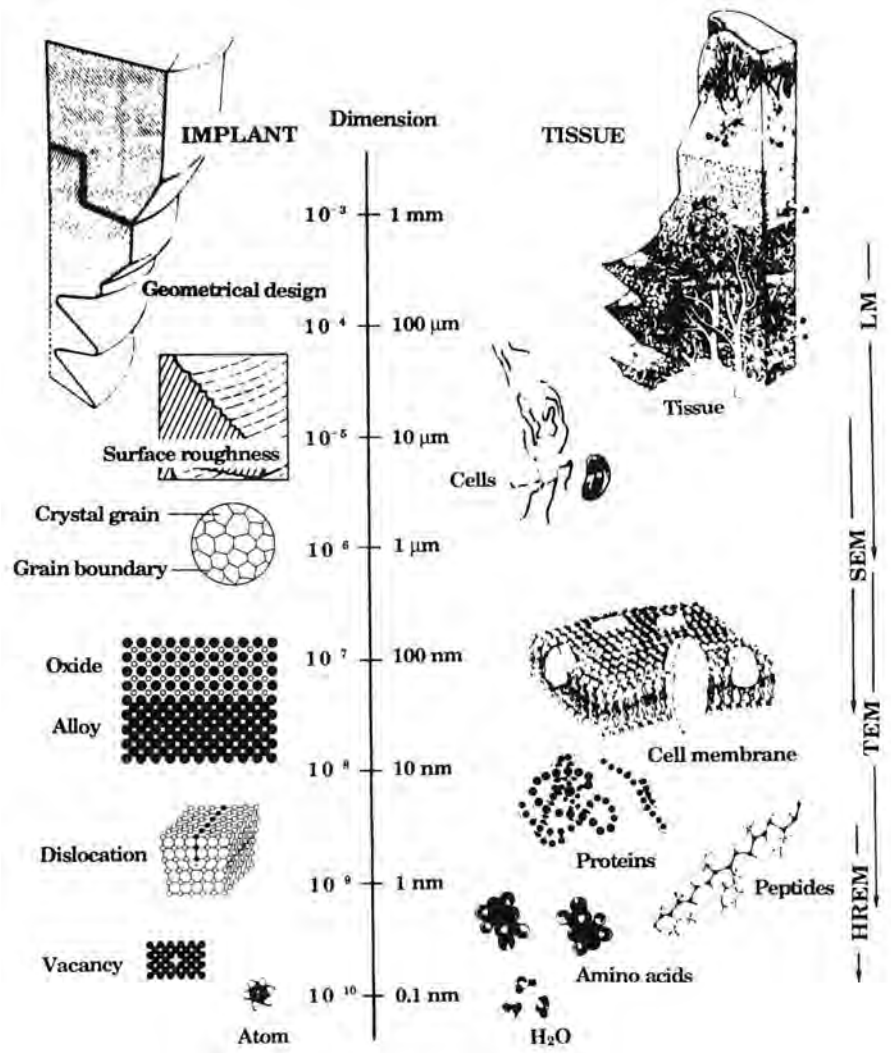


Figure 2. Schematic illustration of the approximate dimensions of some geometrical structures and elements that occur at the interface between a biomaterial and biological system. (After Kasemo and Lausmann, 1988 [2], with permission.)

10.2 Biocompatibility

The term of biocompatibility is defined by Williams [3] as the ability of a material to perform a specific function in the human body with an appropriate host response in the tissues. When a biomaterial is applied to a patient and is in contact with the tissues, the following series of events will be induced from the biomaterial to tissue: (i) an initial physicochemical interaction between constituents of the biomaterial and molecular components of the tissues at the interface; (ii) the generation of the local host response due to the constituents

of the biomaterial immediately adjacent to the interface or to the vicinity of it, and (iii) the systemic distribution and responses following the transporting of physicochemical interaction products away from the reaction sites to remote sites in the body [4].

The initial physicochemical interactions will occur at a solid–liquid interface established between the biomaterial and the body fluids such as blood, serum and extracellular fluid immediately after the implantation. Then, proteins contained in the body fluids will be adsorbed onto the surface of the biomaterial. Adsorbed proteins are affected strongly on cellular response to the biomaterial, for example, adsorbed insoluble globulin enhances the adhesion of fibroblasts, the foreign particles originated from the biomaterial were ingested by macrophages, and thrombus formation on a tube which is fabricated by polyvinyl chloride gave rise on the surface exposed to blood [5]. But also, deterioration of the biomaterial is induced during the interaction at the interface by the mechanisms of depolymerization, cross-linking, oxidative degradation, hydrolysis, crazing and stress cracking in polymers, oxidation, crevice, pitting and galvanic corrosion, stress corrosion cracking and fatigue in metals and alloys, and dissolution in ceramics.

According to a hypothetical model proposed by Williams [4], a sequence of normal wound healing is described chronologically in relation to the tissue responses as follows: (i) an area of wound is filled with blood clot after injury of soft tissue; (ii) blood vessels are dilated, and acute inflammatory response takes place with cells; (iii) the repair process is brought about by capillaries growing into the wound area, by collagen storage in fibroblast to form the connective tissue; and (iv) the healing of the wound is completed by resolution of the fibrous scar tissue. The duration of each stage in the sequence, however, may depend on a balance between the injury and repair, even though no infection occurs. Furthermore, if the damage of the tissue is mild, the inflammatory response may develop into the chronic inflammation. Consequently, the tissue response may never be resolved and may always be accompanied with inflammatory cells as well as new collagen and blood vessels.

By contrast, a sequence of wound healing in the presence of an implant which is not totally inert is described as an interaction between degradation of the biomaterial and host tissue responses as follows: (i) a fibrous capsule is formed surrounding the implant, if the material is not decomposed in the environment; (ii) an extensive response is induced in the tissue, if a release of constituents occurs from the reaction products on the surface of the material; (iii) a chronic inflammatory response takes place in the tissue, if the material leads to persistent stimulus to inflammation by more severe degradation in the tissue; and (iv) an extensive chronic response is provided by aggressive tissue environment, if the degradation is intensified by a cellular activity on the surface of the material.

The remote or systemic effects are explained that if an interfacial reaction leads to the formation of reaction products, and to release of constituents from the products, these constituents are soluble and are taken up rapidly in the extracellular fluids. Therefore, these constituents are able to transport away in either lymphatic or vascular systems to various organs and tissues throughout

the body. Consequently, the solubilized constituents of material are metabolized and excreted, or stored in some tissue depots. As to the fate of these products in the tissue, if these constituents show an irritation to tissues, effects will arise systemically or at remote sites in the body.

Degradation of a biomaterial is an important factor in relation to biofunctionality. If a polymer is degraded relatively rapidly in the body, it can be used as adhesive for sutures, drug-delivering devices and reabsorbable implants serving as temporary scaffolding. Metals and ceramics are applied in orthopedics, because of their relative stability in the body. The degradation of these materials, however, has often led to malfunctioning of prosthetic devices due to the biological medium as well as cellular activity.

Thus, the biocompatibility of candidate materials has been evaluated mainly on the basis of morphological features of the 'fibrous capsule' enveloping the implant which is affected by each stage of the events using light microscopy (LM), scanning electron microscopy (SEM), and transmission electron microscopy (TEM), depending on required dimensional scale.

10.3 Applications of Microscopy to Evaluate Biocompatibility

10.3.1 Biological Responses Between Cultured Cells and Biomaterial

Generally, experiments in vivo do not allow the examination of the specific cellular reaction to the biomaterial, because numerous populations and chemical factors are involved in the implantation. To distinguish and to analyze the sequence of events which have been described in the previous section, and parameters affecting the interaction process, comprehensive studies have been carried out on cell culture in the presence of biomaterials. For example, the biomaterial–bone interface is studied in the relatively well-defined environment of tissue culture such as organ cultures with calvaria and cell culture with bone marrow cells [6].

To assess the influence of synthetic calcium phosphate powders on osteoblast activities, the ability of osteoblasts to migrate from bone explant to other substrata was examined by means of LM, SEM and TEM on isolated human bone cells in culture [7]. Pure β-tricalcium phosphate (β-TCP) and hydroxyapatite (HAP) which have been used as prosthetic and restorative materials for the skeletal system, were powdered in a mortar and passed through a sieve of less than 50 μm. These powder specimens were seeded in the medium after cells in culture reached confluence. After 600 ks (7 days) maintaining the co-culture of cell and material, the internalization of particles of β-TCP or HAP by the two osteoblast cell populations was examined to assess the biocompatibility using SEM and TEM observations. The isolated human bone cells showed capable of ingesting synthetic calcium phosphate granules. It was suggested that the bone cells were able to participate in the degradation of calcium phosphates by the acidic environment of the phagolysosomes which were confirmed by TEM observations.

This technique also enables us to assess cytotoxicity, due to release of cytotoxic products from cell–material interaction (i.e., enzymatic actions). To evaluate cytotoxicity of collagen which is used as an artificial skin, van Luyn et al. [8] developed a sensitive in vitro test system using methylcellulose as the culture medium. This system can be examined by interactions between cross-linked sheep dermal collagen or its extracts and human skin fibroblasts with the methylcellulose culture for a test period of 500 or 600 ks (6 or 7 days) without refreshing the culture medium. Possible cytotoxic effects are measured by counting cell numbers and evaluated cell morphology by phase-contrast light microscopy [9]. Figs. 3a and b exhibit the phase-contrast light microscopic images of a multilayer of completely spread human skin fibroblasts in control culture and of human skin fibroblasts cultured with a disc specimen after 600 ks (7 days), respectively. A significant decrease of cell number and morphological changes of cell are observed in the micrographs. They [10] reported a discrimination of primary cytotoxicity due to extractive constituents from the material, and secondary cytotoxicity induced by a release of cytotoxic products resulting from enzymatic cell–biomaterial interactions, because a pendant molecule can be released by enzymatic attack.

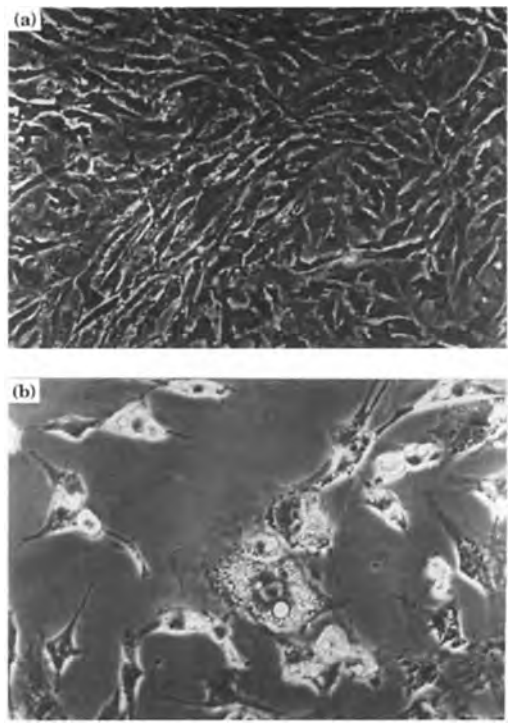


Figure 3. Light micrographs of cell cultures, (a) a multilayer of completely spread human skin fibroblasts in control culture for 0.6 Ms. Some cells contain very small biorefringent particles. (b) Human skin fibroblasts culture with disk-shaped specimen of hexamethylenediisocyanate for 0.6 Ms. Original magnification ×720. (After van Luyn et al., 1992 [9], with permission.)

To elucidate the interaction between tissues and alloy particles which were derived from wear or corrosion of a prosthesis at the cellular level, Rae [11] performed a cell cultural study. His method was as follows. First, foetal rat knee joints were grown in organ culture for 1.2 Ms (2 weeks). At the beginning of the culture period, particulated cobalt, chromium, molybdenum, cobalt–chromium alloy, nickel, or titanium was injected into the joint cavity of limbs. Then the cultures were removed at various time intervals, examined histologically and compared with control cultures. His observation showed that the particulated cobalt, nickel and cobalt–chromium alloy were more toxic than chromium, molybdenum, and titanium. Furthermore, it was found that the synovium was partially destroyed and the metallic particles were enclosed by aggregation of necrotic tissue. The test using monolayer cultures of mouse

peritoneal macrophages showed that particulate titanium, molybdenum and chromium were tolerated by cartilage and synovial cells, and did not impede phagocytosis by macrophages. By contrast, cobalt, nickel and a cobalt–chromium alloy induced localized areas of necrosis in the cartilage and synovium of the knee joint, and damage in the cell membrane of macrophages in the culture.

These methods for assessment by tissue cultures of suitable cells have been developed as a more sensitive technique to evaluate cytotoxicity of candidate materials. The evaluation is made of cell growth rate or cell lysis in comparison to positive and negative controls. Although the assessment by cell culture methods is highly sensitive, this may not accurately reflect the course of events realized at tissue sites in the human body, especially concerning long term effects. There are differences of opinion as to the correlation among the cell culture test, the implantation test into an animal and clinical usage test [12]. Discussion of the biological tests on dental materials has been undertaken by a joint working group for the FDI (Fédération Dentaire Internationale), ADA (American Dental Association) and ISO (International Standards Organization) in 1977. Thus, methods and criteria for the biological tests of candidate materials have been established precisely on dental materials [13].

10.3.2 Biological Responses Between Tissue and Biomaterial

To examine biocompatibility of material by tissue response, experimental works have usually been performed on healthy animals of the same strain in carefully controlled environments. Such an animal model is advantageous to distinguish the complex clinical situation from different well defined elements which can be controlled, so allowing the study of isolated problems. Various species of animal have been employed in such experiments according to the requirements of each study. For example, dogs are often used for studies on orthopedic implant fixation because their bone structure is close to that of humans.

In the preparation of specimens for LM and TEM observation, there is a well-defined series of steps in the fixation, dehydration and embedding of biological materials and, in general, it is only necessary to adopt a very standard routine which is now followed in many laboratories. Furthermore, it is necessary to prepare ultrathin sections of specimens by cutting because of the limited penetration of the incident electron beam in the electron microscope operating at a conventional accelerating voltage up to 100 kV. For adequate penetration and resolution the specimen should not be thicker than 100 nm, since biological specimens already possess a great wealth of structural detail in the 2–100 nm range. The artifacts associated with the cutting are common, because distortions in a fine network structure of the specimens cannot be relaxed after the cutting. An important requirement is the staining of a thin section to provide a sufficient contrast for a differentiation of each part in a microtexture of a biological specimen which is quite different from a crystalline specimen.

Bakker et al. [14] studied phagocyte/polymer interface interactions which will

induce degradation of biomaterials, with dense implants made of polysiloxane and porous implants made of polyether urethane, polypropylene oxide and a polyethylene oxide hydantoin/polytetramethylene terephthalate segmented polyether polyester copolymer for tympanic membrane as a total alloplastic middle ear prosthesis. The implants, which were 100–125 μm thick, were placed at three different sites in the middle ears, in submucosa of tympanic membrane tissue, and between bulla bone and surrounding muscle tissue. The middle ears were dissected from the sacrificed rats after prescribed implantation periods. Sections cut 7 μm in thickness were stained with hematotoxylin-eosin for LM after finishing the standard routine procedure in biological specimens, whereas ultrathin sections were sliced by a microtome, stained with uranyl acetate and lead citrate for TEM. The phagocyte/polymer interactions were evaluated by a composition of the tissue surrounding an implant by measuring of the area in cross section occupied by phagocytes, fibrous tissue and bone using LM as a function of implantation period. They found that the volume of the phagocyte exuded into tissue surrounding the implants, the degradation and fragmentation of the implant. The thickness of fibrous capsules was in relation to the implantation site which was an important factor to assess the performance of implant.

Therefore Bakker et al. [15] also studied the biocompatibility of these elastomers during a *Staphylococcus aureus*-induced middle ear infection which is the most common microorganism causing wound infection in hospitals, because the infection will interfere with the tissue/implant interface and with implant fixation by fibrous tissue and bone, as well as implant degradation. A total number of 983 LM sections were prepared by the standard routine procedure. To assess the biocompatibility, a composition of the tissue surrounding the implant was quantified by measuring the area in cross sections occupied by phagocytes, fibrous tissues and bone as a function of implantation period. The most significant changes attributable to infection were an increase in the fraction of tissue occupied by phagocytes and a decrease in the amount of fibrous tissue with the presence of an implant. The TEM observations of phagocyte/polysiloxane interface showed none of the features associated with implant degradation. This suggested that the polysiloxane elastomer was not subjected to degradation within a one-year implantation period. In addition to the above, significant bacterial degradation or enhanced polymer degradation associated with the inflammatory process was not found in their study.

Collagen-based biomaterials are thought to be the most advantageous biomaterials because the collagen is the main protein component of body tissues such as bones, tendon, skin, ligaments and cartilage, therefore they have been used clinically in bioprosthetic heart valves, vascular prostheses, wound and burn dressings, etc. Thus, application of the collagen as a biomaterial enables 'same with same' replacement implants. The variety of collagens is determined by the structure and number of repeat units in the molecule which will affect the physical properties of collagens, and is classified as types I–IX. Type I collagen is the most plentiful in connective tissue, tendon

and skin. Types II and III exist abundantly in cartilage and blood vessels, respectively [16]. Although commercial products of collagen-based biomaterials are available, modified treatment is necessary to obtain a suitable application in each case as a biomaterial. For instance, cross-linking is able to obtain a strong and stable network of the membrane. A cross-linking agent, however, may influence greatly the immunogenicity of the product. The most widely used cross-linking technique, which is glutaraldehyde treatment, induces a reduction of antigenicity and degradation of the implant.

Vardaxis et al. [17] applied a confocal laser scanning microscopy (CLSM) to study proliferation and differentiation of mesenchymal cells, which are important in the healing process, using an immunostaining technique. They prepared collagen membranes by conventional cross-linking using glutaraldehyde and by a new method involving microwaving and glutaraldehyde. Disk-shaped collagen membranes were implanted in subcutaneous pockets of the ether-anesthetized rats. These membranes were removed from the sacrificed rats after prescribed periods of implantation. The removed membranes and their surrounding tissue were prepared into specimens according to the standard procedure in biological science. Then, the sections were stained by several different procedures to differentiate specified microstructures; for example, a two-step immunostaining procedure was employed with monoclonal antibodies against alpha smooth muscle actin and vimentin labeled with rabbit-antimouse fluorescein isothiocyanate for quantification and three-dimensional imaging by CLSM. They indicated that significant proliferation of

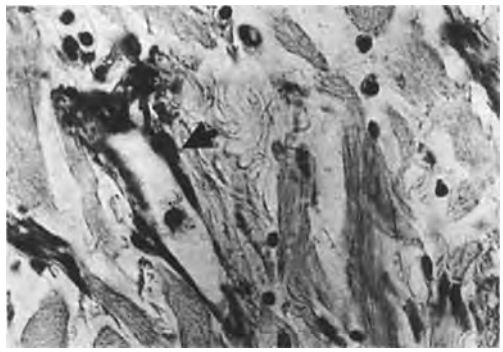


Figure 4. Light microscope of a noncross-linked collagen membrane implanted in animal for 2.8 Ms. An arrow indicates blood vessels formed newly at the rough side of the membrane (400×). (After Vardaxis et al., 1994 [17], with permission.)

small vessels was observed in the interstitium of the fat surrounding the implant, and the endothelial and muscle cells were the proliferating cell nuclear positive after implantation for 350 ks (4 days). The appearance of the cross-linked membrane showed the formation of a new collagen. A prominent vascularization of the implant with vessel ingrowth at the position indicated by an arrow in Fig. 4. Examination by CLSM showed that the tissue reaction around the membrane was one of acute inflammation and early granular tissue formation which is a step in the healing process. The formation of new vessels was also confirmed by CLSM as seen at the position indicated by an arrow in Fig. 5. The geometry of the vessels was followed in the optical serial sections and in the three-dimensional images formed by CLSM.

Scanning tunneling microscopy (STM) has the unique capability of probing the surface structure up to atomic resolution in vacuo, air or liquid. To characterize some aspects of the interface between metallic implants and living tissue, STM

Figure 5. Confocal microscopy image of the connective tissue surrounding a collagen membrane treated with microwave method implanted in an animal for 0.3 Ms. White area shows the membrane. A new blood vessel bud is growing toward the membrane as is indicated by an arrow. (Extended focus image, autofluorescence. Scalebar: 50 μm.) (After Vardaxis et al., 1994 [17], with permission.)

was employed by Emch et al. [18]. They studied biological assays on fibronectin adsorbed on titanium and vanadium coverslips which have been chosen because vanadium is toxic while titanium is inert. Fibronectin is a multifunctional glycoprotein found in both connective tissue and blood, and plays an important role in promoting adherence of staphylococci which cause infection of metallic implants in orthopedic surgery. To examine biological objects with an STM, it was important that a high resolution image was needed immediately during acquisition of data so that real and interesting features were recognized. According to them, this was attained using a video converter which digitized the signal coming from the feedback and the scanning units, and this was transformed into a video signal displayed on a conventional video monitor. They reported that the fibronectin was well adsorbed on the surface of Ti and V substrates, even though these were covered with a natural oxide layer of approximately 10 nm on Ti and 5 nm on V, but it was not bioactive on V substrate.

Ceramics used for implants in medicine and dentistry are of three types: (i) oxide ceramics, such as alumina (Al_2O_3) and zirconia (ZrO_2); (ii) calcium-phosphate ceramics, such as hydroxyapatite (HAP) and tricalcium phosphate (TCP); and (iii) calcium-phosphate containing glass and glasses. Application of ceramics as biomaterials is initially based upon their relative inertness in a biological environment compared to that of metals.

Tissue response between hydroxyapatite granules and human periodontium [19], between single crystal sapphire endosteal dental implant and junctional epithelia of dog [20], between β-TCP granules and bone tissue of monkeys [21] were studied morphologically using LM, SEM and TEM. Certain calcium-phosphate ceramics have a chemical composition similar to the mineral phase in bone tissue. The HAP, TCP and bioglass provide a chemical bond with the bone tissue at their interface after implantation. Orly et al. [22] showed chemical changes in HAP implant under a biological environment realized in human tissue and by cell culture method using TEM and IR spectroscopy. They reported that nucleation and growth of tiny crystals of HAP were

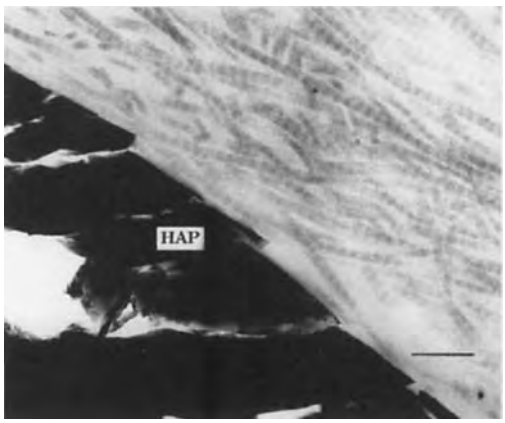


Figure 6. Transmission electron microscope image of collagen fibers running parallel predominantly to the surface of hydroxyapatite. Scalebar: 0.33 μm. (After de Bruijn et al., 1992 [23], with permission.)

found on the surface of the HAP particles which were implanted in human tissue for two months. Initiation of such a mineralization process can be seen in Fig. 6 where collagen fibers are formed predominantly parallel to the HAP surface soaked in cultured bone marrow cells [23]. These collagen fibers integrated and increased in size at the mineralization sites. To perform a stereological examination of osteogenesis at a dental implant, Steflik et al. [24] employed a high voltage electron microscope (HVEM) and conventional TEM under accelerating voltage of 1000 and 100 kV, respectively, because a support by the mandibular or maxillary bone is an important factor for the endosteal implant to succeed. Although densely mineralized collagen fiber matrices were observed in HVEM, they did not give any evidence of mineralization by selected area electron diffraction (SAED) patterns, since the specimens employed were prepared by the routine biological procedure involved in staining. Analysis of the crystalline phase in biologically mineralized tissues is described in Sec. 10.4 of this Chapter.

10.3.3 Interactions Between Blood and Biomaterial

Materials used in the circulatory system (e.g., in artificial heart valves, a totally artificial heart, circulatory-assist devices, oxygenators, liver support systems, renal dialyzers, blood vessel substitutes, and catheters) come into contact with blood flow during service. When blood is contacted with a surface of the artificial organs, various events will be induced at the interface depending upon exposure period. That is, adsorption and/or degradation of blood components, consumption of coagulation factors, and adhesion and aggregation of platelet induce the generation of microthrombi and occasionally life-threatening thromboembolism. The causative mechanisms and prevention of thrombosis and embolization at the interface between blood and biomaterial have been studied by numerous researchers.

Grabowski et al. [25] established a method of evaluating platelet adhesion on various membranes for cardiovascular prostheses exposed to heparinized flowing blood on different species of animal, including human, using SEM. In their study, specimens of membrane were exposed to nonrecirculated blood or to circulated blood of each species for prescribed periods at a constant surface shear rate. The membrane specimens were examined by either phase contrast microscopy or LM after fixation, staining and mounting between a slide glass and cover slip.

SEM observation was also performed. They intended to develop the in situ continuous monitoring of the blood/biomaterial interface reactions by means of video microscopy. Haycox et al. [26] also employed epifluorescent video microscopy to obtain real-time, dynamic information on platelets adhering to candidate biomaterials. Subsequently, Haycox et al. [27], however, stated questions of equivalent reactivity between humans and other animal species. They emphasized the necessity of establishment with respect to a meaningful in vitro blood test compatibility for predicting the performance of candidate material in vivo. To examine platelet interactions with biomaterials, they measured β-thromboglobulin release from platelets into bulk using fresh human blood and a modified Chandler loop system which consisted of a peristaltic pump and a circular tube. The candidate polymers were polyethylene, polyurethane copolymer which consisted of alternating blocks of 'hard' and 'soft' segments, polyvinyl alcohol and a polyurethane prepared with octadecyl pendant groups. The polymer solutions were then solvent cast on the luminal surface of the oxidized polyethylene tube. Adherent platelets were observed using SEM and a colorimetric stain specific for human platelet GPIIIa. They observed four blood reactions: (i) platelets continually adhered, activated and released; (ii) platelets initially adhered and released, but then spread over the surface and passivated it; (iii) platelets did not adhere to the surface but continued to react with unpassivated material; and (iv) platelets were neither adhered nor activated. These reactions were, according to the authors, essential clues to establishing blood compatibility.

10.4 Imaging and Characterization of Tissue Components in the Body

Collagen molecules, which are approximately 300 nm in length, form collagen fibrils by spontaneously aggregating into an ordered arrangement. This ordered arrangement results in a characteristic 60–70 nm transverse D periodicity in the structure. Baselt et al. [28] observed the D periodicity in microfibrils of type I collagen in rat tail tendon and of reconstituted bovine dermal type I collagen by atomic force microscopy (AFM). They showed the D periodicity consisting of alternating grooves and ridges in microfibrils on large-diameter fibrils of the reconstituted bovine dermal collagen as seen in Fig. 7. The authors mentioned that AFM images only reveal a microfibril structure on large-diameter fibrils, while minor band structures appear in TEM images as the most prominent structure on small-diameter fibrils. This is due to

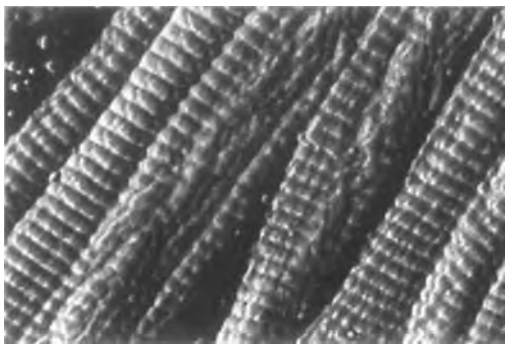


Figure 7. Atomic force microscopic $2.3 \times 1.4\,\mu m$ image of reconstituted microfibrils of bovine dermal collagen. The specimen was fixed in 2.7% glutaraldehyde for 0.6 ks. (After Baselt et al., 1993 [28], with permission.)

an imaging limitation of the AFM. AFM observation, however, is advantageous when examining, without any specimen preparation, cryoprotectants, stainings, embedments, or coatings. It is also possible to make examinations under water by this method.

Dental caries is a multifactorial disease that is a progressive destruction of enamel, dentine and cementum initiated by microbial activity at the tooth surface, even though this explanation is oversimplified. The crystalline phase is biologically mineralized tissues, such as bone and enamel, has a structure similar to hydroxyapatite [$Ca_{10}(PO_4)_6(OH)_2$] and fluoroapatite [$Ca_{10}(PO_4)_6F_2$]. High resolution electron microscopy (HREM) has been an invaluable technique in the study of the mineralized tissues, the structure of biological macromolecules, their interactions and arrangements of cells. Featherstone and Nelson [29] showed a HREM image of the stoichiometric composition of a synthesized HAP crystal with its *c*-axis oriented parallel to the [0 0 1] and its *b*-axis oriented parallel to the [0 1 0]. The former exhibited a lattice image with hexagonal symmetry and the latter represented it with orthorhombic symmetry. They confirmed these structures by comparing with corresponding projections of the apatite structure and with computer-simulated lattice images.

Thereafter, Brès et al. [30] examined irradiation damage of human enamel during observation by HREM at 300 and 400 kV. Their specimens for HREM were sliced using an ultramicrotome, without decalcification and staining, and HREM images were taken parallel to various orientations. They reported that orientation-dependent damage was in the form of mass loss from voids or uniform destruction of crystal structure on the basis of their superb images and analysis including computer image simulation and optical diffraction. A beam-induced diffusion of Ca and O atoms created CaO crystal outgrowths on the enamel surface and crystal patches inside the bulk crystal.

Periodontal disease is caused directly or indirectly by microorganisms in dental plaque. Bacterial growth in the plaque will take up some phosphate from the saliva which is saturated with respect to hydroxyapatite; hence, calcium phosphate is deposited on the teeth surface in the form of calculus. Furthermore, it is thought that the mineralizing process will be accelerated by nucleation of an organic material on the surface of the tooth. Hayashi [31] showed an initial stage of crystallization of calculus on the surface of human enamel by HREM. Ultrathin sections were prepared by cutting with an ultramicrotome after extraction of a human mandibular impacted third molar after extraction and fixation. Because HREM observation was performed with unstained ultrathin sections, it was possible to analyze the HREM images coupled directly with SAED patterns. His observation revealed that calculus, identified as hydroxyapatite needle-like crystals from the SAED patterns, were nucleated coherently on the surface of enamel crystal. Therefore, these crystals of calculus were rigidly formed on the enamel surface and were grown by delivery of constituents from saliva.

Magnetic resonance imaging (MRI) has been currently and widely applied in clinical medicine and dentistry to obtain important diagnostic information, for example, high quality cross-sectional

images of the body using magnetic fields and radiofrequency waves. Lockhart et al. [32] have attempted an in vitro study of an application of three-dimensional MRI to the diagnosis of pulpal pathosis, because dental radiographs which are the most popular technique in dentistry provide two-dimensional imaging only. They used extracted human teeth for MRI after taking radiographs to examine the extent of caries involvement. Multislice imaging allowed for the collection of eight cross-sectional slices at 1.0 mm intervals in coronal and four in apical directions from a central point. The results obtained showed the outlined shape of a tooth including periodontal membrane and internal form of pulp chamber with a high degree of accuracy, thus suggesting the future possibility for accurate diagnosis and treatments of endodontic and periodontal diseases. For example, Fig. 8 shows a cross-sectional image of a molar with MRI taken by Baumann et al. [33]. A texture within the pulp tissue can be seen clearly. Figure 9 exhibits a three-dimensional image of a molar. This suggests that the MRI is able to reproduce a three-dimensional whole configuration of a tooth and a cross-sectional shape of pulpal chamber with a high degree of accuracy using computerized image processing.

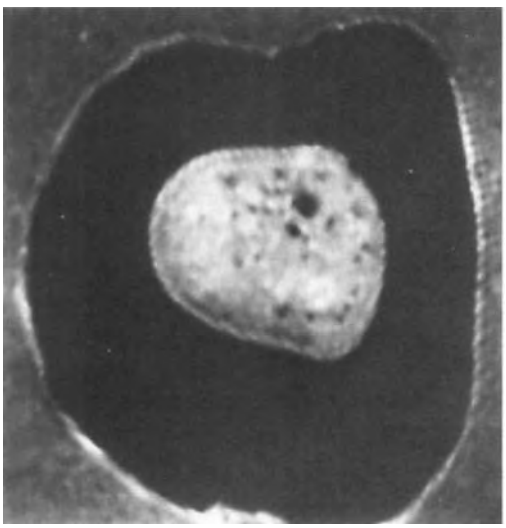


Figure 8. Magnetic resonance microscopic image of a cross-section of the human molar. A texture of pulp tissue can be seen clearly. (After Baumann et al., 1993 [33], with permission.)

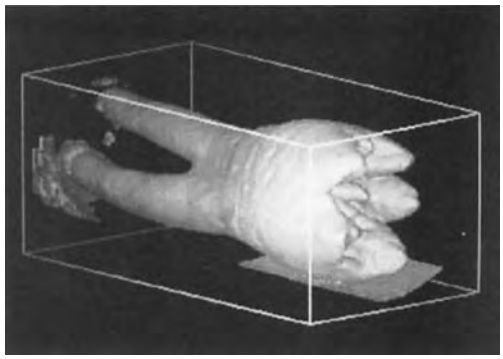


Figure 9. Magnetic resonance microscopic three-dimensional image of the human molar. (After Baumann et al., 1993 [33], with permission.)

Both scanning acoustic microscope (SAM) and scanning laser acoustic microscope (SLAM) have been employed to provide qualitative and quantitative information concerning the material being tested nondestructively. The qualitative information simply provides the morphology in nature and map of the distribution of changes in elasticity and density of a hard tissue on a microscopic scale. Katz and Meunier [34] showed a detailed structure of both human and canine single osteons and osteonic lamellae at a resolution down to approximately 1.7 μm which was well within the thickness of a lamella. When the SLAM is operated in an interference mode, it is possible to obtain the specimen's elastic properties on a

microscopic scale (i.e., quantitative information).

In MRI, SAM and SLAM, the specimens are not necessarily prepared according to the biological treatments such as dehydration, fixation and chemical staining and, therefore, artifacts are not necessarily received. This is thought to be the most advantageous detail compared with other imaging techniques except, of course, for high resolution which is required to distinguish neighboring object points.

10.5 Applications of Microscopy for Characterization of Biomaterials

10.5.1 Characterization of Electrochemical Events on the Surface of Metallic Biomaterials

Surface characterization of metallic biomaterials is particularly important, because surface interactions with surrounding tissue and body fluids lead to various biological and electrochemical events in relation to the biocompatibility and degradation of the implanted devices. The biological events are induced by the release of constituents from corrosion products formed on the surface of a metallic implant in the body. This release of cations is governed as a function of the corrosion rate or oxidation rate of the constituents and of the solubility of the oxide, or first formed corrosion products on the surface of the implant. It is necessary, therefore, to examine the corrosion layer and those factors affecting the corrosion process, so that reactions at the interface between metallic implant and tissue are capable of being minimized, hence reducing inflammation, pain and other problems experienced by the patient.

Electrochemical reactions are examined using anodic polarization, linear polarization and impedance spectroscopy as an averaged measurement because the reactions are detected across an entire surface. However, localized electrochemical events such as etching, selective dissolution, stress corrosion cracking, pitting corrosion, and fretting corrosion processes can be analyzed using surface analytical techniques, for example, Auger electron spectroscopy (AES), X-ray photoemission spectroscopy (XPS or ESCA), electron probe microanalysis (EPMA). To examine morphological aspects on the surface, LM and SEM are commonly used.

It is well known that titanium and its alloys have been used with considerable success for clinical implants because of its excellent corrosion resistance as well as desirable mechanical properties. Occasionally, however, titanium compounds are formed in tissue adjacent to the titanium implants. To elucidate a release mechanism of constituents from an implant, characterization of the passive films on titanium and Ti–6Al–4V alloy was performed in the physiological environment realized by Ringer's solution with changes in pH and amino acid additives using AES and replica TEM by Solar et al. [35]. They showed that the electrochemical reaction was generated by growth of the passive film. They also indicated that the effective electrochemical area of each specimen was less than its geometrical area and was

dependent on both pH and surface finishing conditions.

Bianco et al. [36] studied changes in surface topography and local electronic structure of titanium thin films before and after immersion in an in vitro modeling solution which was a modified Hanks' balanced salt solution with 1.5 mM EDTA. Interactions at the interface between titanium and the solution was observed using STM coupled with tunneling spectroscopy (TS). The topography of the immersed thin film of titanium showed little change from the nonimmersed film specimen, whereas spatial change of the local electronic structure was conspicuous. From the experimental results, they suggested that interactions at the interface between titanium and the solution were not induced uniformly on the surface. They supposed that the change was due to the result of the replacement of the hydroxyl group found on the hydroxylated oxide by phosphate groups.

Stainless steel is the most common implant material in orthopedics. Surgical implants are usually required to possess strength levels far greater than that of bone and have an elastic modulus similar to that of bone. Stainless steel, however, is frequently corroded in the human body. Sundgren et al. [37] examined interface characteristics between stainless steel implants and human oral tissue (both hard and soft) using AES with SEM. They found that the thickness and nature of the passivated oxide layer on the implants depended on the location of the implant sites in the oral region. When the implant was placed in hard tissue, thickness of the oxide layer was not affected or increased, and the oxide was incorporated with calcium and phosphorus. By contrast, when the implant located in soft tissue, the thickness of the oxide layer decreased. These changes were interpreted in terms of oxygen pressures and metabolic activity.

Although comprehensive studies on the corrosion resistance of commercial dental gold alloys have been performed, it is not easy to understand their corrosion characteristics completely because these alloys have extremely complex combinations of constituents, frequently containing five or more elements. In these alloys, the overall corrosion resistance is thought to be controlled by the total amount of precious metals (i.e., nobility). Nakagawa et al. [38] examined the effect of microstructure on the corrosion of Au–Cu–Pd and Au–Cu–Ag ternary and Au–Cu–Ag–Pd quaternary alloys using anodic potentiodynamic and potentiostatic polarization tests in deaerated 1% NaCl solution at 310 K. To consummate changes in microstructure associated with phase transformations, a single phase of solid solution and of $L1_0$ type ordered structure was obtained by annealing at 1073 K and at 573 K for appropriate periods, then quenched in ice brine, respectively, in Au–Cu–Pd ternary alloys. A two-phase coexisting microstructure, which consisted of Cu-rich and Ag-rich f.c.c. phases, was also realized by a similar heat treatment in Au–Cu–Ag ternary and Au–Cu–Ag–Pd quaternary alloys. The phase transformation and the associated microstructure as well as changes in crystal symmetry were confirmed by TEM and SAED techniques. They proved that the single phase microstructure of Au–Cu–Pd ternary alloys exhibited high corrosion resistance. The high nobility Au–Cu–Ag ternary alloys also showed high corrosion resistance

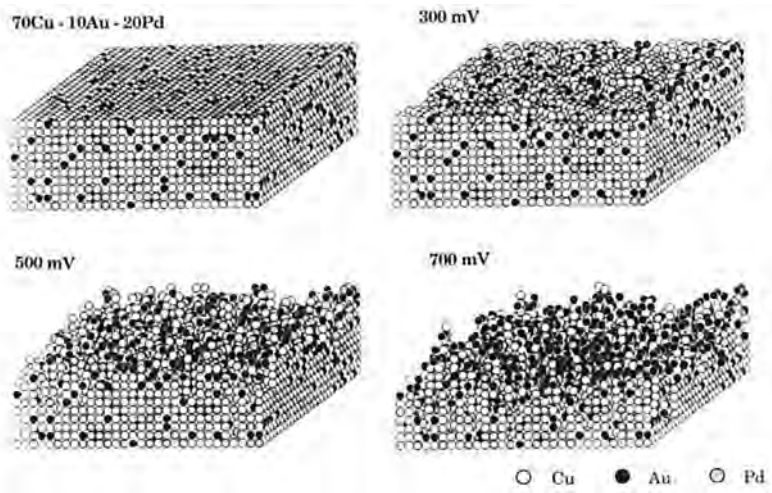


Figure 10. Computer simulated schemes during anodic dissolution of the Cu–10 at.% Au–10 at.% Pd alloy polarized at various potentials. (The figure reproduced newly for the book with courtesy Nakagawa et al., 1993 [39].)

regardless of its two-phase coexisting microstructure. However, the generalized finding was that the two-phase microstructure had lowered the corrosion resistance compared with the single phase alloys. They also found that anodic potentiodynamic polarization curves of the two-phase microstructure could be explained by superimposing the curves of each phase using the principle of additivity. In fact, the potential at which the current density increased significantly corresponded to that of each phase, and the amount of the current density was calculated as the summation of each phase in proportion to their volume fractions.

In noble metal alloys, the preferential or selective dissolution of less noble elements and surface enrichment with noble elements may take place if the alloy is polarized at a lower potential than that of the anodic reaction of the noble elements, as is well known. Nakagawa et al. [39] tried to explain their experimental results by computer simulations of the anodic dissolution process in Au–Cu–Pd ternary alloys using the dissolution model and taking into account the effect of atomic interaction energy. Dissolution probability of those constituents which exhibited the dissolution rate were defined by considering the standard electrode potential of dissolution reaction and anodic interaction between Cu and the other atoms. Figure 10 shows the changes in the surface configurations with increasing polarization potential (i.e., dissolution probability) in a Cu–10 at.% Au–20 at.% Pd alloy. It is possible to observe that significant dissolution occurs and the surface of the alloy becomes rougher with increasing polarization potential. Thus, the computer simulation qualitatively showed a reproduction of their experimental results of anodic potentiostatic polarization concerning the dependence of the anodic dissolution behavior on the polarization potential and alloy composition, and the atomic concentration depth profile of the constituents after anodic polarization. These

results suggested that atomic interaction among the constituents of alloys was an important factor governing the anodic dissolution features of alloys. It is thought that STM and AFM observation is necessary to confirm this dissolution model.

Recently, Gilbert et al. [40] have introduced a new technique capable of imaging highly localized electrochemical currents associated with charge transfer reactions on the surface of metallic biomaterials, using a scanning electrochemical microscope (SECM). This is a nonoptical scanning microscope which can be examined in two different modes: the oxygen reaction rate on the surface of a specimen mode and the ion release and deposition mode. They carried out SECM examinations to evaluate electrochemical reactions or cation release processes in Ni minigrids, an intermetallic compound, γ_1, formed in crystals of dental amalgam and Co–Cr–Mo alloy beads sintered to the main body of a hip joint replacement prosthesis which exhibited significant differences in geometries and corrosion processes. They have emphasized that the SECM is a valuable tool for imaging microelectrochemical processes on the surfaces of metallurgical polished alloy specimens and a wide variety of other surfaces of biological interest where charge transfer occurs.

10.5.2 Coherent Phase Diagram of Au–Cu–Ag, Au–Cu–Pt Ternary and $[Au_x(Ag_{0.24}Cu_{0.76})_{1-x}]_{0.958}$–$Pt_{0.015}$ Quaternary Systems

It is well known that most dental gold alloys containing platinum and palladium exhibit conspicuous age-hardening characteristics. In 1905, an age-hardenable gold–copper–silver–platinum alloy, which we now realize from its composition was hardenable by AuCu ordering, was developed and put on the market for the first time by the S. S. White Dental Manufacturing Co. (USA). It is astonishing to note that age-hardening was being applied in dental practice so early on [41].

The essential components of commercial dental gold alloys are thought to be gold, copper and silver. A knowledge of equilibrium phase diagrams is an important aid in predicting the phase transformations resulting from the aging of alloys; the ordering regions in the Au–Cu–Ag ternary system have not been established with sufficient accuracy. A brief description of the historical background on the determination of the Au–Cu–Ag ternary system is given in [42].

Generally, a great deal of experimental work is needed; hence, long periods are needed to allow a complete evolution to the equilibrium state, also there is much expense involved in the case of precious metal alloys to determine the equilibrium phase diagram of the ternary alloy system. Yamauchi et al. [43] calculated a 'coherent' phase diagram of the Cu–Au–Ag ternary system, using the determined interaction energy parameters to fit only the experimental binary phase diagrams of Au–Cu, Cu–Ag and Ag–Au systems. Even though it was intended to represent only 'coherent' equilibria (i.e., when all the possible phases differ from each other merely by the arrangement of the constituent atoms of copper, silver or gold on a fixed f.c.c. lattice of the disordered solid solution) the calculated phase diagram showed striking similarities with the

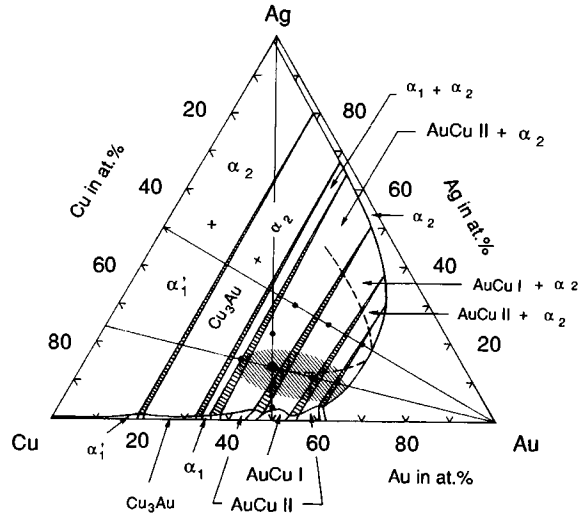


Figure 11. Isothermal section of a plausible coherent phase diagram of the Au–Cu–Ag ternary system at 575 K. (After Hisatsune et al., 1991 [45].)

experimentally derived phase diagram determined by X-ray powder diffraction according to Uzuka et al. [44]. This calculated phase diagram was confirmed to refer to TEM experimental data obtained by Yasuda's group. Figure 11 shows an isothermal section at 573 K of the plausible Au–Cu–Ag ternary phase diagram which was constructed by superimposing the theoretical phase diagram calculated using the cluster variation method [43] on experimental data [45]. In the phase diagram, the shaded area indicates the composition region corresponding to commercial dental gold alloys and the dots show the composition of alloys examined by TEM and HREM.

According to the hypothesis of Allen and Cahn [46], a coherent phase must be metastable and must be unstable in the presence of the incoherent phase; the coherency strain must also be present in the coherent multiphase structure, and dislocation must be visible at the interfaces with the incoherent phase to reduce coherency strain. They showed that the coherent phase diagram of the Fe–Al binary system could be obtained by TEM and SAED, as long as the study satisfied their hypothesis. A coherent phase diagram gives us more effective information to predict phase transformations related to age-hardening in an alloy system than an incoherent equilibrium phase diagram.

Figure 12 is part of the coherent phase diagram of Au_x–$(Ag_{0.24}Cu_{0.76})_{1-x}$ section in the Au–Cu–Ag ternary system which was determined by means of TEM bright- and dark-field imaging coupled with the SAED technique, because this section passes through the composition region of commercial dental gold alloys, as seen in the shaded area of Fig. 11 [47]. The formation region of long period superstructure of AuCuII appears coexisting with a silver-rich disordered f.c.c. α_2 phase. The Cu_3AuII long period superstructure was also found in the $Cu_3Au + \alpha_1 + \alpha_2$ three phases coexisting region. Here, the α_1 phase is a Cu-rich disordered f.c.c. structure. In relation to this system, a coherent

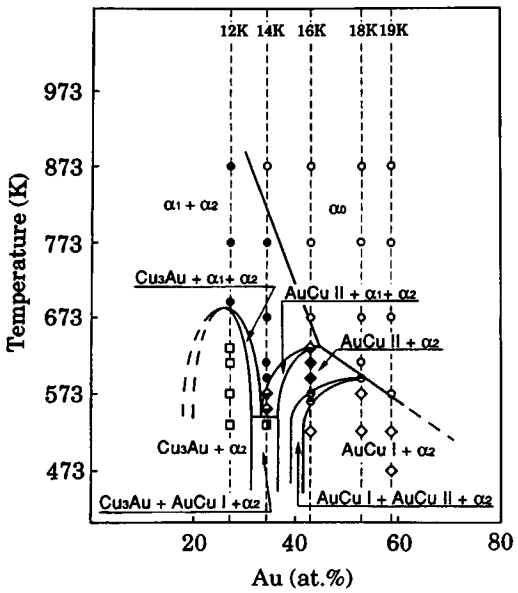


Figure 12. Coherent phase diagram of the Au_x–$(Ag_{0.24}Cu_{0.76})_{1-x}$ pseudobinary section in Au–Cu–Ag ternary system. (After Nakagawa and Yasuda, 1988 [47].)

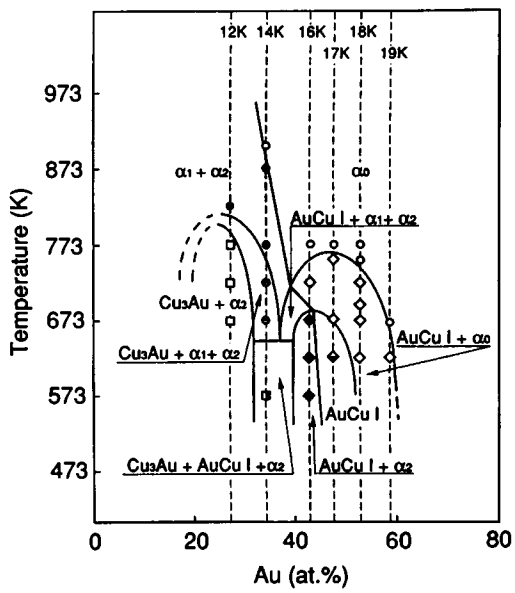


Figure 13. Coherent phase diagram of the $[Au_x$–$(Ag_{0.24}Cu_{0.76})_{1-x}]_{0.985}$–$Pt_{0.015}$ pseudobinary section in Au–Cu–Ag–Pt quaternary system. (After Iwanuma, 1994 [48], with permission by his supervisor.)

phase diagram of more complex quaternary alloys $[Au_x$–$(Ag_{0.24}Cu_{0.76})_{1-x}]_{0.985}$–$Pt_{0.015}$ pseudobinary section in the Au–Cu–Ag–Pt system was also determined recently using TEM and SAED by Iwanuma [48], as shown in Fig. 13 in which any long period AuCuII superstructure cannot be detected. The coherent phase diagram of AuCu–Pt pseudobinary section was determined also using TEM and SAED by El Araby et al. [49]. They showed that the AuCuII ordered region disappeared on the phase diagram in the composition range over approximately 1.3 at.% Pt.

A HREM coupled with SAED is an indispensable technique to identify a minor phase in multiphase structure for the determination of a coherent phase diagram. For instance, Yasuda et al. studied structural aspects of the AuCuII ordered phase associated with order twinning in AuCu–3 at.% Ag [50] AuCu–9 at.% Ag [51] and AuCu–14 at.% Ag [52] alloys of HREM and SAED. The α_0 phase which was predicted from the phase rule could be found adjacent to the tip of the microtwin platelets when they were in contact with other orientation variants of the microtwin platelets as an inevitable consequence of strain accommodation in AuCu–3 at.% Ag alloy [50]. In AuCu–9 at.% Ag alloy, they observed that a characteristic mosaic structure consisted of cuboidal blocks of the α_2 phase enclosed by two different orientation variants of the AuCuII twinned ordered phase by HREM [51]. Their results revealed the atomic arrangement of each phase, the interfacial structure between the coexisting phases on the scale of interatomic distance and strain field induced

by order twinning at corners in the cuboidal α_2 phase.

10.5.3 Structural Aspects Associated with Phase Transformation and Age-hardening Mechanisms in Dental Gold Alloys

In dental gold alloys, age-hardening is induced mainly by AuCu ordering. Microstructural changes related to AuCu ordering and two-phase decomposition were studied by TEM, HREM and SAED in Au–Cu–Ag ternary alloys [50–52]. In Au–Cu–Ag ternary alloys, four distinguishable phase transformations occurred depending on the silver content and on the annealing temperature: (a) the formation of the AuCuI and/or AuCuII long period ordered platelets and its twinning characterized by a stair-step made by a nucleation and growth mechanism (N–G process); (b) AuCuI and/or AuCuII ordering and two-phase decomposition promoted by the N–G process; (c) the ordering resulting from spinodal decomposition; and (d) the formation of the lamellar structure developed from grain boundaries by discontinuous precipitation.

The HREM study showed that process (a) was brought about in a composition located outside of the miscibility gap of two-phase decomposition in Au–Cu, that is, low content Ag alloys in the AuCu–Ag pseudobinary system [50]. Changes in microstructures induced by process (b) were complex; for example, two types of precipitate, spindle type and chain type, were formed by the N–G process upon annealing above the critical temperature for ordering, T_c. Below the T_c, these precipitates changed to a sandwich structure consisting of cuboidal blocks of α_2 phase and AuCuII ordered thin platelets along two of their three faces, and only their lateral faces were not decorated and connected directly with the α_1 matrix. In a low silver content alloy, AuCu–5.6 at.% Ag, the α_2 phase was formed as finely dispersed precipitates, sometimes as small as 10 nm × 2 nm, coexisting with AuCuI and/or AuCuII ordered phase [53]. In process (c), below the spinodal temperature, the modulated structure was generated in the early stage of aging, then the cuboidal blocks of the α_2 phase were completely sandwiched by the AuCuII ordered thin platelets [52].

Although there are four types of phase transformation giving rise to age-hardening, only two types of phase transformation are reported in commercial dental gold alloys. The age-hardening characteristics of commercial dental gold alloys are also studied by Yasuda's group using TEM, SAED, HREM, electrical resistivity measurements, hardness tests and X-ray diffraction. In an Au–21.8 at.% Cu–8.8 at.% Pt–8.8 at.% Pd alloy, age-hardening occurred in two stages. The first stage arose from the nucleation of AuCuI ordered platelets and was characterized by a slow growth rate of the ordered nuclei [54]. In the second stage, simultaneous ordering and precipitation took place. An alternating lamellar structure consisting of AuCuI ordered platelets and platinum-rich f.c.c precipitates was identified by TEM and SAED. Studies for elucidating age-hardening mechanisms on a series of commercial dental gold alloys were performed, since it was expected that the age-hardening mechanism varied greatly

depending upon the concentration of gold in the alloys.

The two alloys studied were Au–26.5 at. % Cu–18.2 at. % Ag–7.9 at. % Pd (16 K/S) and Au–25.3 at. % Cu–27.4 at. % Ag–7.6 at.% Pd (14 K–S) [55, 56]. In these alloys, two stages of AuCuI ordering could be distinguished by TEM and SAED studies of isochronal aging of a supersaturated solid solution. In the first stage, ordering of a metastable AuCuI′ ordered phase formed as a stepwise arrangement of platelets situated on the {1 0 0} planes. Prolongation of the aging period caused the formation of a lamellar structure composed of the equilibrium AuCuI ordered and a silver-rich disordered f.c.c. phases. The metastable AuCuI′ ordered platelets were formed prior to the equilibrium AuCuI ordered phase. The AuCuI′ platelet formation was accompanied, in the 16 K–S alloy, by a contraction of approximately 0.07 along the *c*-axis of the AuCuI′ in the surrounding matrix. In the 14 K–S alloy, the contraction along the *c*-axis was approximately 0.06. The strain introduced by the tetragonality of the AuCuI′ structure led to substantial hardening in the early stages of aging.

Thereafter, a different type of age-hardening mechanism was found in a commercial dental gold alloy, Au–31.7 at.% Cu–8.1 at.% Pd–5.3 at.% Ag (18 K–S) [57]. The X-ray and SAED studies as well as TEM showed that Au_3Cu $L1_2$ type superlattice with thermal antiphase domain boundaries (APBs) was formed by aging at 673 K. It was thought that the hardening arose through the mechanism of antiphase domain size effect. The $L1_0$ type AuCuI superlattice was also formed in regions adjacent to the Au_3Cu ordered phase in the later stages of aging below 623 K. The TEM micrographs showed that the AuCuI ordered platelets were formed in the Au_3Cu ordered matrix along ⟨1 0 0⟩ directions. Thus, it was concluded that the hardening should be attributed not only to the APBs size effect, but also to the elastic strain field induced by the formation of the ordered AuCuI platelets in the aging temperature range below 623 K.

Although the hardening mechanism in dental gold alloys was thought to be attributed to a coherency strain field at the interface between the AuCuI ordered platelets and the surrounding matrix, the configuration of the interface, that provides direct evidence of the presence of the coherency strain fields, could not be deduced by conventional TEM. Therefore, HREM coupled with SAED was employed to elucidate structural configuration of the interface between the AuCuI and Au_3Cu ordered phases in the 18 K–S alloy [58]. An HREM image taken from the specimen aged at 673 K for 1.8 Ms with the incident electron beam parallel to the [1 0 0] direction showed that the square pattern of bright dots which had the same configuration and scale as the projection of the copper columns in Au_3Cu ordered structure. From this configuration it was clearly evident that the copper columns were at the correct positions with respect to the basic f.c.c. lattice (i.e., at each corner of the square). A shift in the arrays of the bright dots observed across the interfaces were composed of the APBs. The presence of APBs introduces a local change in chemical composition by excess atoms with respect to the stoichiometric composition of an ordered phase. Then, these excess atoms will be segregated along the

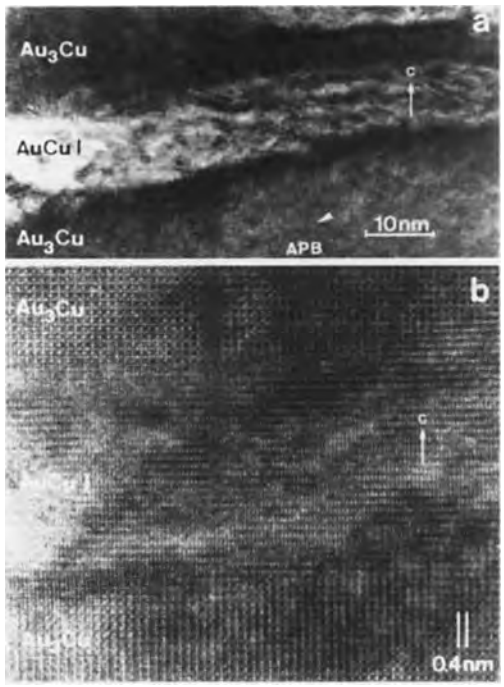


Figure 14. Bright field high resolution electron microscopic images of a commercial dental gold alloy (18 K–S) aged at 573 K for 1.8 Ms. (a) Only the superlattice reflections were selected for imaging; (b) all reflections up to 220_{fcc} were selected for imaging. (After Yasuda et al., 1986 [58].)

APBs. This segregation contributed to the hardening as well as to the higher energy at the APBs. When aging was carried out below 623 K, the AuCuI ordered platelets were formed in regions adjacent to the Au_3Cu ordered matrix. Figure 14a was produced by using 0 0 1, 1 0 0, 1 1 0 and equivalent superlattice reflections, the fundamental reflections being excluded, and contain an interface between the AuCuI platelet and the Au_3Cu ordered matrix of a specimen aged at 573 K for 1.8 Ms. Only one-dimensional resolution is observed in the AuCuI ordered platelet. For the HREM image in Fig. 14b, the f.c.c. reflections were also included in the imaging aperture to show in detail the arrangement

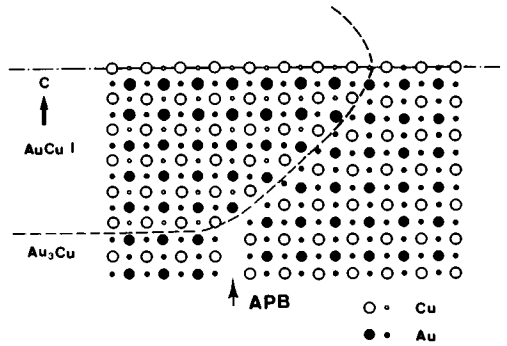


Figure 15. Schematic representation of an AuCu(I) platelet in an Au_3Cu matrix for axial ratio c/a of 0.9; an antiphase domain boundary (APB) is induced automatically if copper continuity is required along the center of the platelet (large circles are at level 0; small circles are at level 1/2). (After Yasuda et al., 1986 [58].)

of the atom columns along the interface between the AuCuI ordered platelet and the Au_3Cu ordered matrix. In spite of the disturbance caused by the presence of strain and strain contrast, the interface is clearly visible due to the small difference in atomic distance between the AuCuI and Au_3Cu ordered phases. It is clear that the amount of elastic strain increases in proportion to the thickness of the AuCuI ordered platelet, that is the length along the c-axis, within the range of the elastic limit.

If, as shown schematically in Fig. 15, it is assumed that at the top of such a platelet, the pure copper (0 0 1) planes in AuCuI coincide with the mixed gold and copper planes in Au_3Cu, then with an axial ratio $c/a = 0.9$ for a platelet with a halfwidth of 5 AuCuI unit cells, the pure copper planes will now coincide with the pure gold planes. For a c/a ratio of 0.945 which was the measured value from the diffraction patterns, the width of the AuCuI platelet will be 18 or 19 unit cells thick. This explanation is in good agreement

with the experimental observations; indeed, such a mismatch of the APB is observed in Fig. 14a as is indicated by the arrow head. Thus, it was concluded that the AuCuI ordered platelets, being coherent with the Au_3Cu ordered phase as matrix, gave rise to a considerable amount of elastic strain which could be a major contribution to the age-hardening in dental gold alloys. It was the first time to date that the atomic configuration in ordered phases induced by aging was produced and analyzed in commercial dental gold alloys.

10.6 Concluding Remarks

This section has reviewed some applications of microscopy which evaluate or characterize materials according to the required scale in the field of biomaterials science. A large number of papers in this field appear annually in the literature, the subject of which varies according to the investigator's background: that is, leanings towards physics, chemistry or biology, as well as clinical evaluation by clinicians in medicine and dentistry. In this article, we classified such contributions into two parts at the interface of an implant and surrounding tissue. In other words, biomaterials science is composed of a biological scientific approach and materials scientific approach, thus making the subject very much an interdisciplinary science. Clearly, from the contents of this contribution, the reader will appreciate that the discussion is selective and represents our special interests. However, from the references cited, the reader will be able to make further excursion into this fascinating domain.

10.7 References

[1] A. B. Ferguson in *Bone and Joint Surgery* (Ed.: C. O. Bechtol, A. B. Ferguson, P. G. Laing), Williams & Wilkins, Baltimore **1959,** pp. 1–18.

[2] B. Kasemo, J. Lausmaa in *Surface Characterization of Biomaterials, Progress in Biomedical Engineering* (Ed.: B. D. Ratner), Elsevier, Amsterdam **1988,** pp. 1–12.

[3] D. F. Williams in *Medical and Dental Materials, Materials Science and Technology, Vol. 14* (Ed.: D. F. Williams), VCH, Weinheim **1992,** pp. 1–27.

[4] D. F. Williams in *Concise Encyclopedia of Medical and Dental Materials* (Ed.: D. F. Williams), Pergamon Press, Oxford **1990,** pp. 51–59.

[5] T. A. Horbett in *Biomaterials: Interfacial Phenomena and Applications, Advances in Chemistry Series 199* (Ed.: S. L. Cooper, N. A. Reppas), American Chemical Society, Washington, DC **1982,** pp. 233–244.

[6] C. Maniatopoulos, J. Sodek, A. H. Melcher, *Cell Tissue Res.* **1988,** *254,* 317–330.

[7] M. Grégoire, I. Orly, J. Menanteau, *J. Biomed. Mater. Res.* **1990,** *24,* 165–177.

[8] M. J. A. van Luyn, P. B. van Wachem, L. Olde Damink, H. Ten Hoopen, J. Feijen, P. Nieuwenhuis, *J. Mater. Sci.: Mater. Med.* **1991,** *2,* 142–148.

[9] M. J. A. van Luyn, P. B. van Wachem, L. Olde Damink, P. J. Dijkstra, J. Feijen, P. Nieuwenhuis, *J. Biomed. Mater. Res.* **1992,** *26,* 1091–1110.

[10] M. J. A. van Luyn, P. B. van Wechem, L. H. H. Olde Damink, P. J. Dijkstra, J. Feijen, P. Nieuwenhuis, *Biomaterials* **1992,** *13,* 1017–1024.

[11] T. Rae in *Biocompatibility of Implant Materials* (Ed.: D. F. Williams), Sector, London **1976,** pp. 55–59.

[12] D. S. Smith in *Biocompatibility of Dental Materials, Characteristics of Dental Tissues and Their Response to Dental Materials, Vol. 1* (Ed.: D. S. Smith, D. F. Williams), CRC Press, Boca Raton **1982,** pp. 1–37.

[13] K. Langeland in *Concise Encyclopedia of Medical and Dental Materials* (Ed.: D. F. Williams), Pergamon Press, Oxford **1990,** pp. 59–69.

[14] D. Bakker, C. A. van Blitterswijk, S. C. Hesseling, J. J. Grote, W. T. Daems, *Biomaterials* **1988,** *9,* 14–23.

[15] D. Bakker, C. A. Blitterswijk, S. C. Hesseling, W. Th. Daems, W. Kuijpers, J. J. Grote, *J. Biomed. Mater. Res.* **1990,** *24,* 670–688.

[16] D. F. Williams in *Concise Encyclopedia of Medical and Dental Materials* (Ed.: D. F. Williams), Pergamon Press, Oxford **1990,** pp. 112–115.

[17] N. J. Vardaxis, J. M. Ruijgrok, D. C. Rietveld, E. M. Marres, M. E. Boon, *J. Biomed. Mater. Res.* **1994,** *28*, 1013–1025.
[18] R. Emch, X. Clivaz, C. Taylor-Denes, P. Vaudaux, P. Descouts, *J. Vac. Sci. Technol.* **1990,** *A8*, 655–658.
[19] I. Orly, B. Kerebel, J. Abjean, M. Heughebaert, I. Barbieux, *Bull. Group Int. Rech. Sci. Stomatol. Odontol.* **1989,** *32*, 79–86.
[20] R. V. McKinney, D. E. Steflik, D. L. Koth, *J. Periodontol.* **1985,** *56*, 579–591.
[21] M. Franch, V. de Pasquale, A. Ruggeri, R. Strocchi, *Bull. Group Int. Rech. Sci. Stomatol. Odontol.* **1991,** *34*, 123–131.
[22] I. Orly, M. Gregoire, J. Menanteau, M. Heughebaert, B. Kerebel, *Calcif. Tissue Int.* **1989,** *45*, 20–26.
[23] J. D. de Bruijn, C. P. A. T. Klein, K. de Groot, C. A. van Blitterswijk, *J. Biomed. Mater. Res.* **1992,** *26*, 1365–1382.
[24] D. E. Steflik, A. L. Sisk, G. R. Parr, L. K. Gardner, P. J. Hanes, F. T. Lake, D. J. Berkery, P. Brewer, *J. Biomed. Mater. Res.* **1993,** *27*, 791–800.
[25] E. F. Grabowski, P. Disheim, J. C. Lewis, J. T. Franta, J. Q. Stropp, *Trans. Am. Soc. Artif. Intern. Organs* **1977,** *23*, 141–151.
[26] C. L. Haycox, B. D. Ratner, T. A. Horbett, *J. Biomed. Mater. Res.* **1991,** *25*, 1317–1320.
[27] C. L. Haycox, B. Ratner, *J. Biomed. Mater. Res.* **1993,** *27*, 1181–1193.
[28] D. Baselt, J.-P. Revel, J. D. Baldeschwieler, *Biophys. J.* **1993,** *65*, 2644–2655.
[29] J. D. B. Featherstone, D. G. A. Nelson, *Scanning Microsc.* **1989,** *3*, 815–828.
[30] E. F. Brès, J. L. Hutchison, B. Senger, J.-C. Vogel, R. M. Frank, *Ultramicroscopy* **1991,** *35* 305–322.
[31] Y. Hayashi, *J. Electron Microsc.* **1993,** *42*, 342–345.
[32] P. B. Lockhart, S. Kim, N. L. Lund, *J. Endodontics* **1992,** *18*, 237–244.
[33] M. A. Baumann, D. Gross, V. Lehmann, K. Zick, *Schweiz Monatsschr. Zahnmed.* **1993,** *103*, 1407–1414.
[34] J. L. Katz, A. Meunier, *J. Biomech. Eng.* **1993,** *115*, 543–548.
[35] R. J. Solar, S. R. Pollack, E. Korostoff, *J. Biomed. Mater. Res.* **1979,** *13*, 217–250.
[36] P. D. Bianco, P. Ducheyne, B. Bonnell, *J. Mater. Sci.; Mater. Med.* **1992,** *3*, 28–32.
[37] J.-E. Sundgren, P. Bodö, I. Lundström, A. Berggren, S. Hellem, *J. Biomed. Mater. Res.* **1985,** *19*, 663–671.
[38] M. Nakagawa, S. Matsuya, M. Ohta, *J. Mater. Sci.; Mater. Med.* **1992,** *3*, 114–118.
[39] M. Nakagawa, T. Shiraishi, S. Matsuya, M. Ohta, *J. Mater. Sci.* **1993,** *28*, 1199–1204.
[40] J. L. Gelbert, S. M. Smith, E. P. Lautenschlager, *J. Biomed. Mater. Res.* **1993,** *27*, 1357–1366.
[41] K. Yasuda, *Gold Bull.* **1987,** *20*, 90–103.
[42] K. Yasuda in *Concise Encyclopedia of Medical and Dental Materials* (Ed.: D. F. Williams), Pergamon Press, Oxford **1990,** pp. 197–205.
[43] H. Yamauchi, H. A. Yoshimatsu, A. R. Forouhi, D. de Fontaine in *Proceedings of the 4th International Precious Metals Conference*, Toronto, Ont., Canada (Ed.: R. O. McGachie, A. G. Bradley), Pergamon Press **1980,** pp. 241–249.
[44] T. Uzuka, Y. Kanzawa, K. Yasuda, *J. Dent. Res.* **1981,** *60*, 883–889.
[45] K. Hisatsune, K.-I. Udoh, B.-I. Sosrosoedirdjo, T. Tani, K. Yasuda, *J. Alloys & Compounds* **1991,** *176*, 346–254.
[46] S. M. Allen, J. W. Cahn, *Acta Metall.* **1975,** *23*, 1017–1026.
[47] M. Nakagawa, K. Yasuda, *J. Less-Comm. Met.* **1988,** *138*, 95–106.
[48] K. Iwanuma, *Phase transformation related to age-hardening in* $[Au_x-(Ag_{0.24}Cu_{0.76})_{1-x}]_{0.985}-Pt_{0.015}$ *system*, PhD Thesis, Nagasaki University Graduate School of Dental Science **1994,** pp. 1–148.
[49] A. M. El Araby, Y. Tanaka, K.-I. Udoh, K. Hisatsune, K. Yasuda, *J. Alloys & Compounds* **1994,** *206*, 217–224.
[50] K. Yasuda, M. Nakagawa, K.-I. Udoh, G. Van Tendeloo, J. Van Landuyt, *J. Less-Comm. Met.* **1990,** *158*, 301–309.
[51] K. Yasuda, K. Hisatsune, K.-I. Udoh, Y. Tanaka, G. Van Tendeloo, J. Van Landuyt, *Dentistry in Japan* **1992,** *29*, 91–102.
[52] K.-I. Udoh, K. Yasuda, G. Van Tendeloo, J. Van Landuyt, *J. Alloys & Compounds* **1991,** *176*, 147–158.
[53] K. Yasuda, M. Nakagawa, G. Van Tendeloo, S. Amelinckx, *J. Less-Comm. Met.* **1987,** *135*, 169–183.
[54] K. Yasuda, M. Ohta, *J. Less-Comm. Met.* **1980,** *70*, 75–87.
[55] K. Yasuda, K.-I. Udoh, K. Hisatsune, M. Ohta, *Dent. Mater. J.* **1983,** *2*, 48–58.
[56] K.-I. Udoh, K. Hisatsune, K. Yasuda, M. Ohta, *Dent. Mater. J.* **1984,** *3*, 253–261.
[57] K.-I Udoh, K. Yasuda, M. Ohta, *J. Less-Comm. Met.* **1986,** *118*, 249–259.
[58] K. Yasuda, G. Van Tendeloo, J. Van Landuyt, S. Amelinckx, *J. Dent. Res.* **1986,** *65*, 1179–1185.

Handbook of Microscopy
S. Amelinckx, D. van Dyck, J. van Landuyt, G. van Tendeloo

11 Carbon

11.1 Introduction

In view of its industrial importance in general, and in particular in the field of nuclear energy, carbon has been the subject of a large amount of research applying a wide range of physico-chemical methods. Among these microscopy holds a prominent place, since the morphology, microstructure, and texture are important parameters for the behavior of carbon, mainly as graphite.

We shall limit ourselves in this contribution to the description of a number of representative features observed by means of various microscopy methods, also illustrating the power of these techniques in this particular field of research.

11.2 The Different Forms of Carbon

11.2.1 Graphite

Elemental carbon occurs in different solid forms. The low temperature, low pressure phase is graphite (see, for example, Marsh [1]); its structure consists of planar hexagonal honeycomb networks of sp_2-bonded carbon atoms (Fig. 1a). The stacking of the layers in normal hexagonal graphite ($a = 0.2456\,\text{nm}$, $c = 0.6696\,\text{nm}$) can be represented by the sequence ABAB.... However, a rhombohedral variety, represented by the sequence ABCABC... occasionally occurs, in particular in heavily deformed graphite. Synthetic graphites are often orientationally disordered (turbostratic); the material cleaves very easily along the basal planes, since the bonding between the layers is of the weak Van der Waals type.

11.2.2 Diamond

The high temperature, high pressure form of carbon is cubic diamond ($a = 0.3567\,\text{nm}$). This structure consists of a spatial network of tetrahedra formed by sp_3-bonded carbon atoms (Fig. 1b), forming an f.c.c. lattice, the stacking sequence in terms of (111) layers being ...AαBβCγ.... Also a hexagonal form (Londsdaleite), having a wurtzite-type structure with stacking sequence ...AαBβ... is formed under special circumstances (e.g., in meteorite craters) [2]. In the past, diamond has been synthesized under high pressure and temperature in the presence of various catalysts (for a review,

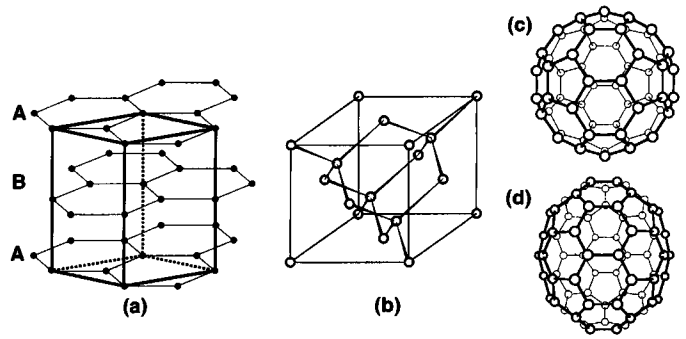


Figure 1. The different forms of solid elemental carbon: (a) graphite: sp_2-bonding, (b) diamond: sp_3-bonding, (c) fullerenes: C_{60} and (d) C_{70}.

see, e.g., Wilks and Wilks [3]). More recently, low pressure methods based on plasma-assisted chemical vapor deposition (CVD) have been developed, allowing the deposition of thin films of diamond on a crystalline substrate (for example, silicon) (see, e.g., references [4–7] and [3] for a review). The microstructure of these films is remarkable and important for the applications concerned.

11.2.3 Fullerenes

As well as these two forms of elemental carbon, a third form was discovered recently. Under certain circumstances, carbon has been found to form closed cage-like molecules [8], which give rise to predominantly Van der Waals bonded molecular crystals [9]. Such molecules are found in the soot produced in an electric arc between graphite electrodes under the reduced pressure of helium [9]. The most abundant species in the soot are C_{60} and C_{70}. Models of two of the simplest among these globular molecules, called fullerenes, are shown in Figs. 1c and d. They can be extracted from the soot by organic solvents and can be separated by chemical methods such as liquid phase chromatography, and then purified by selective sublimation [10]. C_{60} and C_{70} were studied initially because they were most readily available in appreciable quantities.

Owing to the initially very limited availability of pure, crystalline C_{60} and C_{70}, the structure and microstructure of these molecular crystals were initially studied almost exclusively by means of electron microscopy [9, 11–14] and X-ray diffraction [9, 15–17], and to a lesser extent by means of atomic force and scanning tunneling microscopy (especially for the observations of thin films) [18–24], whereas scanning electron microscopy was used to study the morphology (Fig. 2)

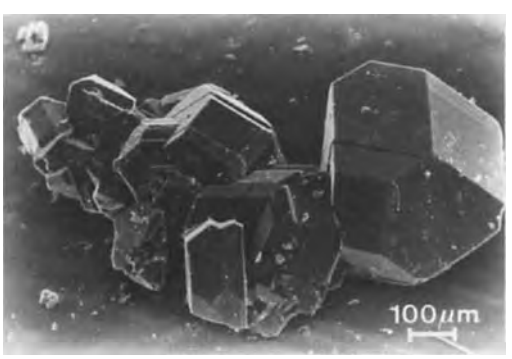


Figure 2. Scanning electron microscopic (SEM) image of sublimation-grown C_{60} single crystals (Van Tendeloo et al. [14]).

(for example, references [14, 25]). The crystal structures and phase transitions upon heating or cooling of fullerene crystals are reviewed in Fisher and Heiney [26].

The molecule C_{60} is the most symmetrical and most spherical fullerene; it is a globular cage consisting of 20 hexagonal and 12 pentagonal edge-shearing meshes. Each pentagon is surrounded by five hexagons and each hexagon by three pentagons and three hexagons (Fig. 1c). All closed-cage fullerenes contain a mixture of hexagons and pentagons; they satisfy Euler's rule for closed polyhedra (Appendix A, Sec. 11.7.1 of this Chapter). The number of pentagons is invariably 12 because it is the 'closure condition', i.e., twelve is the number required to cause a total convex curvature of 4π (Appendix B, Sec. 11.7.2 of this Chapter). The C_{70} molecule is an elongated, ellipsoidal cage differing from C_{60} by the presence of one more 'belt' of hexagons (Fig. 1d). Whereas the molecular point group symmetry of C_{60} is I_h, that of C_{70} is only D_{5h}.

11.2.4 Nanotubes

As a by-product of fullerene production by the electric arc method, it was recently discovered by means of transmission electron microscopy (TEM) that carbon often forms fine tubular fibers, with diameters of the order of nanometers and lengths of up to micrometers [27]. Such fibers have potential applications and have therefore been studied extensively (reviewed in references [28–30]). Owing to their small size, individual fibers can only be observed conveniently by electron diffraction and electron microscopy.

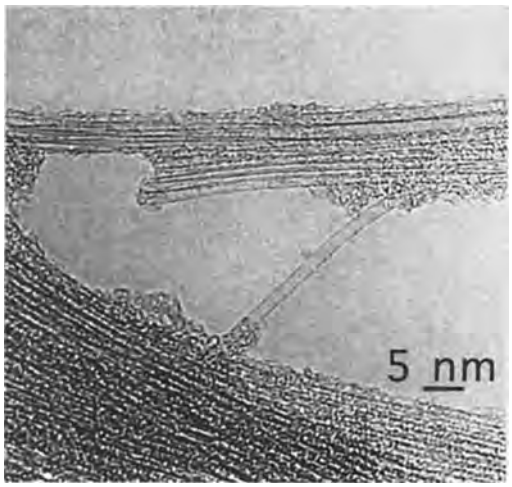


Figure 3. A single wall carbon tube (TEM); some bundles of parallel, single wall tubes are also present.

These fibers are either concentric cylinders of graphene sheets [27] or multiturn scrolls made of a single graphene sheet or a mixture of both [31, 32]. Single layer cylindrical fibers (Fig. 3) have recently been produced via a catalytic method, introducing the transition metal catalyst in the graphite electrodes of the electric arc [33–35]. Even more recently, high yields of single sheet tubes have been obtained by laser evaporation in an oven at 1200°C of a carbon target containing a small amount of transition metal atoms such as nickel or cobalt [36, 37]. Without a catalyst, the arc-grown tubules are usually multilayered and often contain imperfections, a number of which will be discussed below.

Under adequately chosen conditions, helix-shaped carbon fibers are produced via the catalytic decomposition at relatively low temperatures (~600–800°C) of hydrocarbons (Fig. 4) [38, 39]. Under particular conditions, conically wound, cylindrically shaped solid needles of carbon can be produced (Fig. 5) [40, 41].

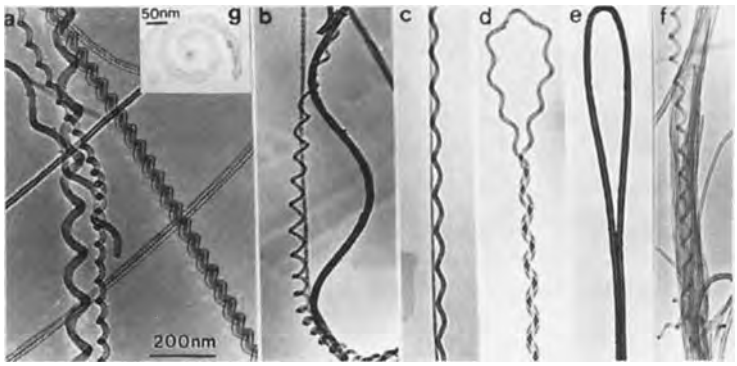


Figure 4. (a) to (g) Helix-shaped carbon tubules produced by the catalytic method and observed by TEM. Note the variety of shapes (Amelinckx et al. [212]).

11.3 Graphite

11.3.1 Specimen Preparation

Specimens with a uniform thickness over large areas and suitable for TEM diffraction contrast studies can easily be prepared by the repeated cleavage of single crystals (as occur in the Ticonderoga limestone, USA). Pieces of adhesive tape are hereby stretched over flat glass microscopy slides in order to keep the tape flat, and subsequently pressed with the adhesive coating against both sides of a graphite flake. The slides are repeatedly pulled apart avoiding shearing. Very thin flakes then adhere to the tape. Dissolving the glue in an organic solvent liberates these flakes, which then float in the liquid and can be deposited on standard grids.

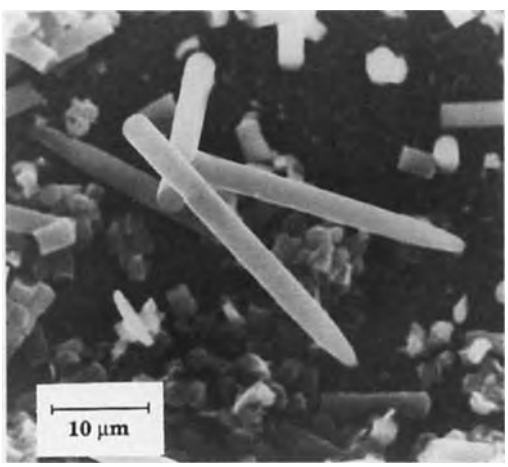


Figure 5. SEM image of conically wound graphite needles produced from CO on an SiC substrate (Amelinckx et al. [40]).

11.3.2 Transmission Electron Microscopy (TEM) Studies

Owing to pronounced cleavage, specimens are invariably parallel to the basal plane. Since the glide planes of dislocations are also parallel to the (0001) plane, the geometry of the specimen lends itself particularly well to a study of basal plane dislocation configurations. A number of these are described in Chap. 4.1, Sec. 1.1.18 of Vol. 1 of this Handbook and in more detail in Amelinckx et al. [42]. Basal plane glide dislocations are always dissociated in two partials of the Shockley type with Burgers vectors $A\sigma$ and σB, according to the equation $AB = A\sigma + \sigma B$, using the notation illustrated in Fig. 6.

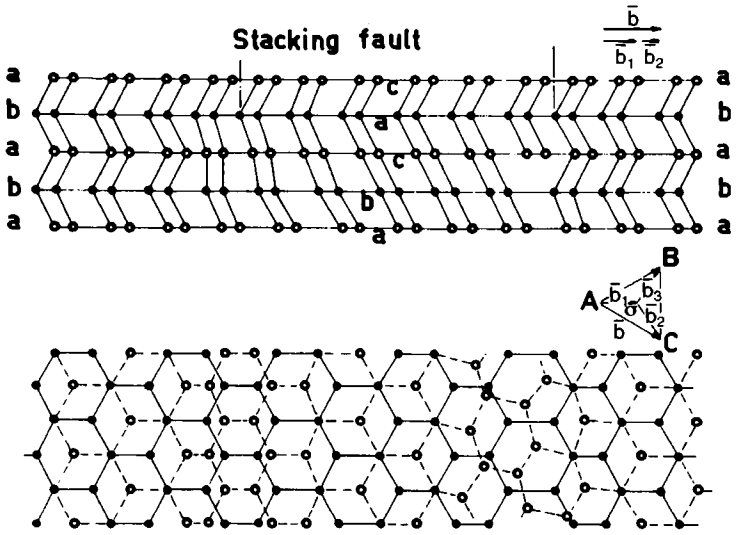


Figure 6. Schematic cross-sectional and plane view of a glide dislocation ribbon in graphite. A ribbon of intrinsic stacking fault separates the partial dislocations. A reference triangle for denoting Burgers vectors is shown as well (Amelinckx et al. [42]).

Imaging under two beam conditions with a $30\bar{3}0$ type active reflection reveals the three families of partials with Burgers vector $\pm A\sigma$, $\pm B\sigma$, and $\pm C\sigma$ (Fig. 7). These conditions were used to image single dislocation ribbons and study the variation of their width d with the angle ϕ enclosed by the direction of the total Burgers vector AB of the dislocation pair and the direction of the ribbon (Fig. 8). From

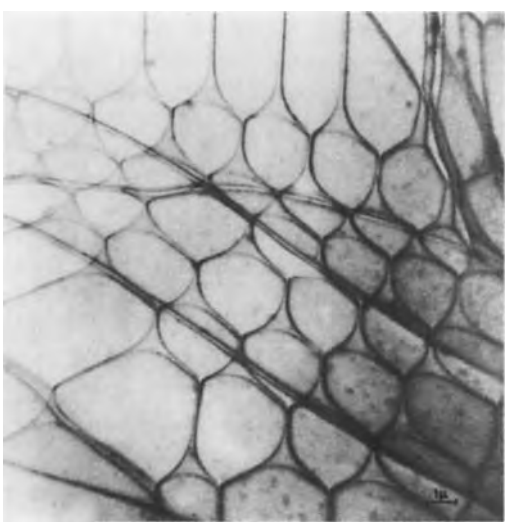


Figure 7. Network of partial dislocations in the basal plane of graphite observed in two-beam bright field diffraction contrast (TEM) (Amelinckx and Delavignette [44]).

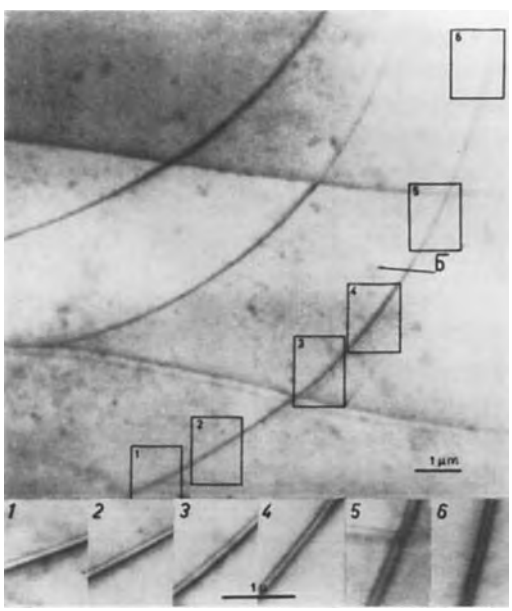


Figure 8. Curved single ribbon of partials in graphite. Note the variation in ribbon width with the orientation of the ribbon (TEM) (Delavignette and Amelinckx [43]).

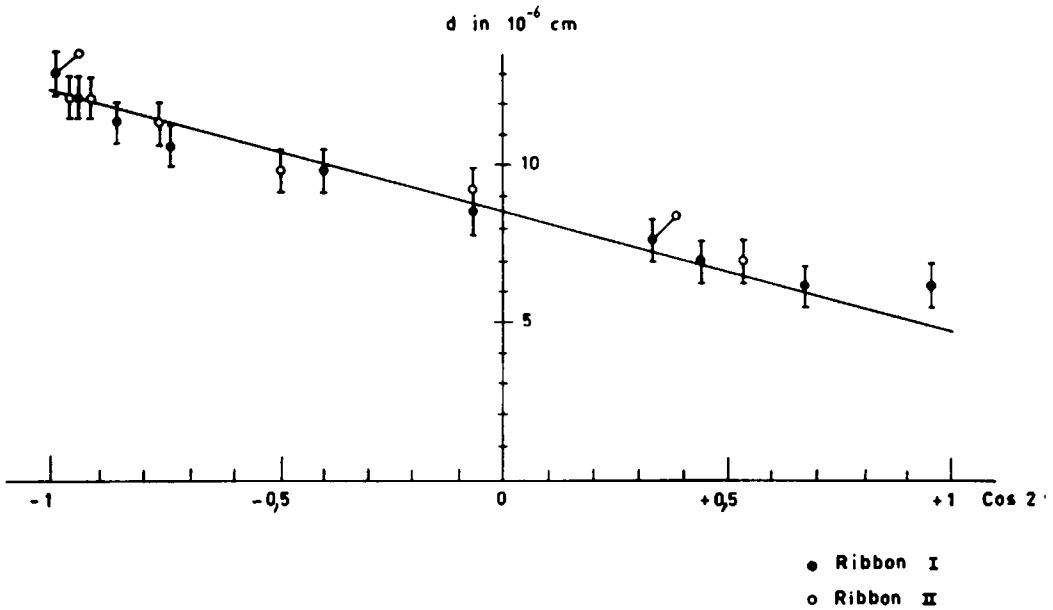


Figure 9. Plot of the ribbon width d as a function of a character of the dislocation ribbon: $\cos\phi$. According to Eqs. (1) and (2), the graph is a straight line of which the slope yields ν and the intercept d_0. This allows us to compute γ/μ from Eq. (2) (Amelinckx and Delavignette [43]).

this variation the stacking fault energy γ can be deduced, as well as the effective value of Poisson's ratio ν, as was shown for the first time in graphite [43]. This method is based on the equations

$$d = d_0\left[1 - \frac{2\nu}{(2-\nu)}\cos 2\phi\right] \qquad (1)$$

$$\gamma = \frac{\mu \boldsymbol{b}^2}{8\pi d_0}\left(\frac{2-\nu}{1-\nu}\right) \qquad (2)$$

(Fig. 9), where μ is the shear modulus and $\boldsymbol{b}$ the Burgers vector of the partial dislocations. From the linear plot of d versus $\cos 2\phi$, such as in Fig. 9, d_0 is deduced from the intercept and ν from the slope; γ/μ then follows from Eq. (2). The underlying theory is found in Amelinckx et al. [42]. It should be noted that to deduce the real ribbon width from the image the sideways image shift should be taken into account, which can either be in the same sense for both partials, or in the opposite sense; in the latter case it can be either inwards or outwards, leading respectively to an apparently smaller or larger width (see Chap. 4.1, Sec. 1.1.18 of Vol. 1 of this Handbook).

The Burgers vector of individual partials is determined using the image extinction criterion $\boldsymbol{g}\cdot\boldsymbol{b} = 0$. Partial dislocations go out of contrast for reflections of the type $11\bar{2}0$, i.e., for reflections of the 'second' hexagon. The 'absent' partial dislocation has a Burgers vector that is perpendicular to the active diffraction vector (Fig. 10b). One image extinction is sufficient, since in the particular case of graphite the Burgers vectors of glide dislocations must be parallel to the basal plane [44].

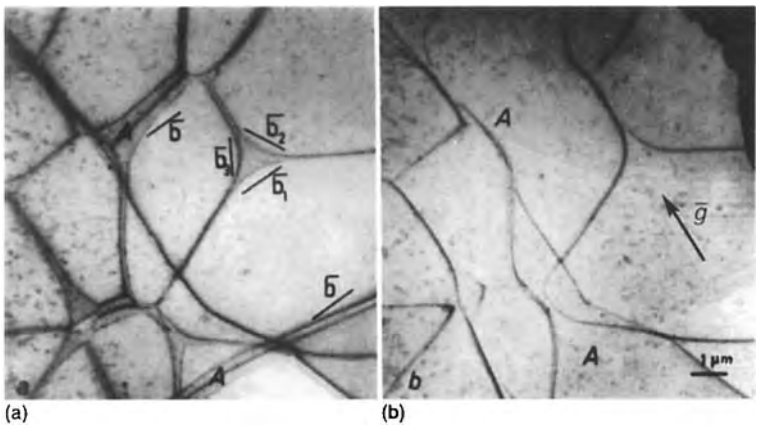


Figure 10. Extinction of partial dislocations in a network situated in the basal plane of graphite. Note the simultaneous extinction of the three partials in the triple ribbons at A (TEM). (a) All dislocations in contrast. (b) One set of dislocations is extinct (Amelinckx and Delavignette [44]).

Imaging in a $10\bar{1}0$ type reflection (i.e., in the first hexagon) either in bright field (BF) or in dark field (DF) reveals the stacking fault areas, as in Fig. 10a, as darker areas. This figure reveals two features of particular interest. The extended dislocation node, consisting of three partial dislocations forming a triangular area limited by curved partial dislocations, provides an alternative method to deduce the stacking fault energy from the smallest radius of curvature of the partials. The theoretical basis of this measurement is discussed in the literature [45, 46].

The second feature is the 'triple ribbon' in A, consisting of three parallel partials separated by two strips of stacking fault [43, 44]. The simultaneous extinction of the three partials in the triple ribbon of Fig. 10b leads to the conclusion that the three partials have the same Burgers vector, which is indicated in Fig. 10a [44]. This does not exclude the fact that they have different line images (Part IV, Chap. 1, Sec. 1.1.18 of Vol. 1 of this Handbook). Models for such triple ribbons, which were first discovered in graphite, are discussed in Part IV, Chap. 1, Sec. 1.1.20 of Vol. 1 of this Handbook and Amelinckx et al. [42]. They are very sensitive probes for measuring the stacking fault energy since they are wider than single ribbons. They also make it possible to observe the effect of surface steps, such as XY, on the ribbon width, as is evident in Fig. 11. The proximity of a surface decreases the repulsive force between the partials without affecting the magnitude of the stacking fault energy, which determines the attractive interaction; as a result, the total width decreases as the ribbon approaches the surface. On determining the stacking fault energy from ribbon widths, the widest ribbon should therefore always be used.

Graphite foils often contain twist boundaries situated in the (0 0 0 1) plane [42]; these possibly result from deformation during specimen preparation, but they must be frequent in 'turbostratic' graphite. They consist of a network of extended and contracted nodes (Fig. 12) formed by the intersection of two families of ribbons with different total Burgers vectors. In Fig. 12b the faulted areas, which are on the concave side of the partials, show up brighter, which is rather exceptional since this requires the faulted area to be close to a

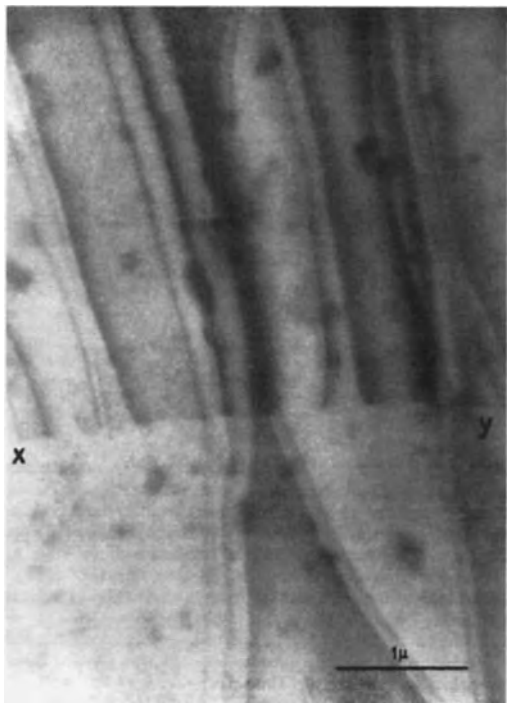


Figure 11. Triple ribbon in the (0001) plane of graphite passing under a surface step XY, hereby changing its width (TEM) (Amelinckx et al. [42]).

surface (see Part IV, Chap. 1, Sec. 1.1.18 of Vol. 1 of this Handbook). The partial dislocations in such a network have a predominantly screw character in agreement with the twist character of the boundary.

In graphite heated in vacuum to about 3000°C and subsequently quenched, circular faulted loops can be observed in the basal plane (Fig. 13) (see, for example, Amelinckx et al. [42]). They are attributed to the precipitation of vacancies into disks followed by collapse of the adjacent lattice planes. The Burgers vectors of these loops were shown to be inclined with respect to the basal plane, showing that on collapsing a lateral offset takes place. Also, heavily neutron-irradiated and subsequently annealed graphite was shown to contain loops of a similar type using the method described in Part IV, Chap. 1, Sec. 1.1.18 of Vol. 1 of this Handbook.

In bromine-intercalated graphite the boundary between the intercalated and the pristine regions is formed by a

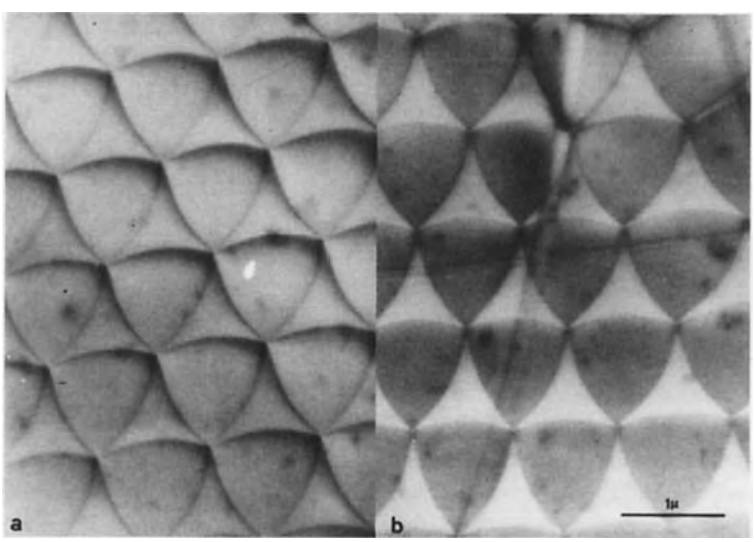


Figure 12. Network of contracted and extended nodes in graphite under two different two-beam diffraction conditions. In (a) line contrast, and in (b) the faulted areas show up bright because the network is close to the surface (TEM) (Amelinckx et al. [42]).

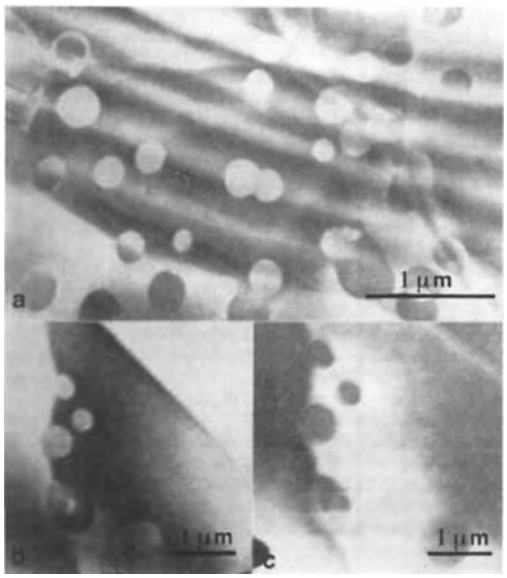


Figure 13. (a) Circular dislocation loops in the (0 0 0 1) plane, formed from quenched-in vacancies in graphite. Note the fault contrast inside the loops (TEM). The fault contrast is inverted between (b) and (c), which represent the same area (Amelinckx et al. [42]).

particular type of dislocation which can be revealed in TEM and of which the motion can be studied in situ [47, 48]. It was shown by TEM that the bromine enters the graphite preferentially along stacking faults.

11.3.3 Optical Microscopy

Conventional optical reflection microscopy and interference microscopy (so-called Fizeau fringes [49]) have been used to study the surface oxidation of graphite by oxidizing agents (e.g., O_2) (see Thomas [50] for a review). The surface is found to oxidize by the lateral removal of C-layers nucleated at preferential sites which are interpreted as the emergence points of nonbasal dislocations [50]. Even single vacancies are held potentially responsible for the nucleation of pits [50]. The effect of metallic catalyst particles on the oxidation behavior was also studied (see, e.g., Hennig [51, 52] and Thomas [50] for a review). Nonbasal dislocations have been revealed by etching and studied by optical microscopy of (0 0 0 1) surfaces in references [53, 54].

11.3.4 Scanning Tunneling Microscopy

Scanning tunneling microscopy (STM) allows us to reveal the surface of graphite at the atomic level (Fig. 14). It turns out that, against intuitive expectation, the honeycomb structure is not revealed but the image is instead a close-packed hexagonal array of dots with lattice parameter a [55]. This is interpreted by assuming that the geometry of the two

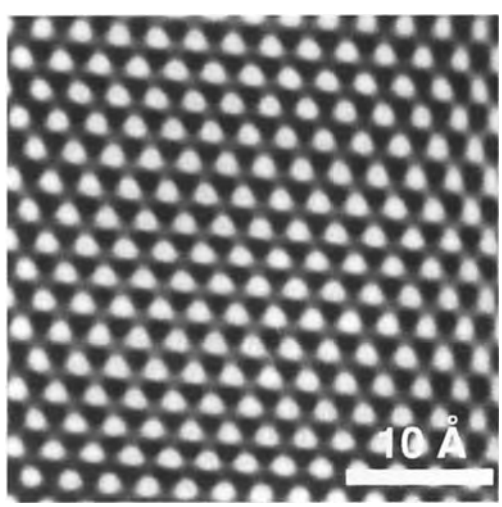


Figure 14. STM image of the (0 0 0 1) surface of graphite. Only half of the surface atoms are imaged (Kelty and Lieber [58]).

uppermost layers is revealed. Two types of graphite surface atoms have then to be considered. Half of the surface atom has a neighbor at $c/2$ along the c-direction, the other half has not; the first neighbor along the c-direction is at a distance c. This is obvious from the atomic structure of graphite (Fig. 1a). It appears that only one kind of surface atom is actually revealed (Fig. 14) [56]. Furthermore, large vertical corrugations have been observed on STM images of graphite surfaces in air. It was argued that this effect could be attributed to elastic interactions between the graphite surface and the STM tip through a layer of contamination [57]. It was found that images of clean graphite surfaces in vacuum do not show such corrugations.

Stage 1 potassium-intercalated graphite (KC_8) has been imaged by means of STM [58]. A $2a$ superlattice was detected. The presence of the potassium ions causes the graphite layers to become stacked A–A instead of A–B as in pristine graphite; nevertheless, not all carbon sites show up as identical peaks. Elastic deformations of the graphite surface are not held responsible for the observed period doubling in the image; this is attributed to electronic effects associated with the ordering of the potassium ions (Fig. 15).

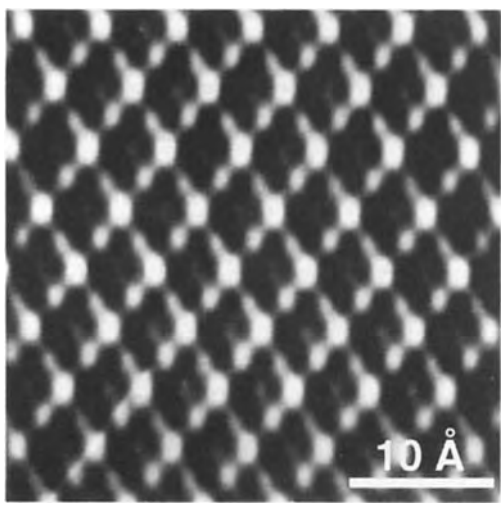


Figure 15. STM image of (0 0 0 1) surface of graphite intercalated with potassium (C_8K); note the period doubling in the image (Kelty and Lieber [58]).

11.4 Diamond

11.4.1 Natural Diamond

The microscopic characterization of diamond as a gem is one of the subjects treated in Chap. 1.7 of this Volume; in the following paragraph we treat the materials science aspects.

Natural diamonds have been studied extensively by means of optical surface methods by Tolansky and co-workers (see, for example, reference [59]). They discovered the presence of 'trigons' and shallow triangular pits on natural (1 1 1) faces (Fig. 16). These were interpreted by a number of authors as etch figures [60], and by others as growth features [61]. The methods used were phase contrast reflection optical microscopy (Fig. 16a) and multiple beam interferometry in monochromatic light (Fig. 16b) [49, 61]. Fizeau-type fringes (Fig. 16b) were produced using an interferometer consisting of an optical flat pressed against the surface to be studied, both being silvered to about 80–90% transmission. In this manner, surface steps of only a few nanometers high can be detected and measured optically by multiple beam interferometry, but the lateral resolution

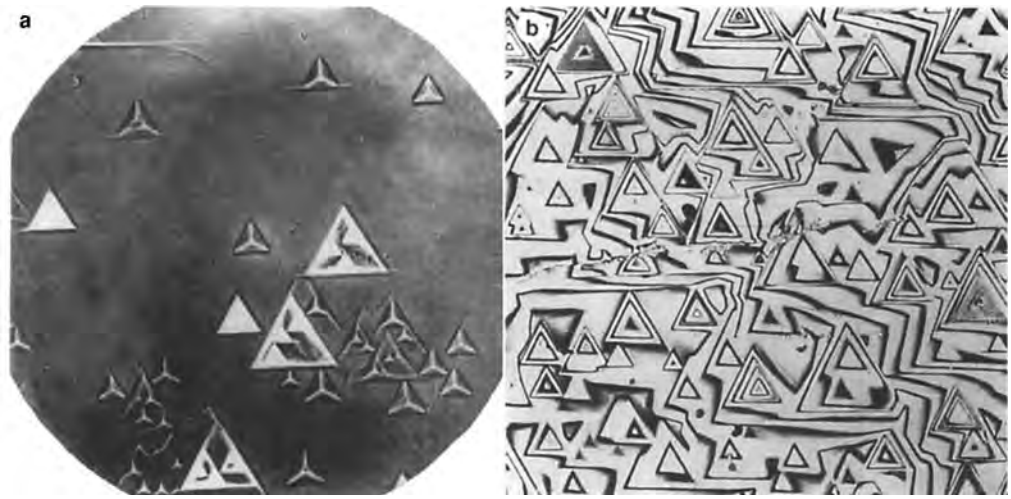


Figure 16. Trigons in diamond on a (1 1 1) face, (a) imaged by means of phase contrast reflection optical microscopy, and (b) by means of multiple beam interferometry (Tolansky [61]).

is of course limited to that typical of optical microscopes (Part I, Chap. 1 of Vol. 1 of this Handbook). The same techniques were used to study the anisotropy of the abrasion rate along different directions on several crystal faces (see, e.g., Wilks and Wilks [62], or Wilks and Wilks [3] for an overview), a problem of technical importance.

X-ray topography has been used extensively to study dislocations by Lang and coworkers (for overviews, see Part II, Chap. 5 of Vol. 1 of this Handbook and references [63, 64]). Owing to their high degree of perfection, natural diamonds make appropriate samples for X-ray topography, which is a low resolution method (Fig. 17). Natural diamond was found to contain very few dislocations (Fig. 18).

Diffraction contrast TEM has also been used in the study of dislocations in natural diamond; in particular the weak beam method was applied to study the splitting in partial dislocations and to derive the stacking fault energy [65, 66].

TEM studies led to the discovery of so-called nitrogen platelets (Fig. 19) [67–71], which are essentially Guinier–Preston zone-like precipitates containing layers of nitrogen atoms, which in dispersed form

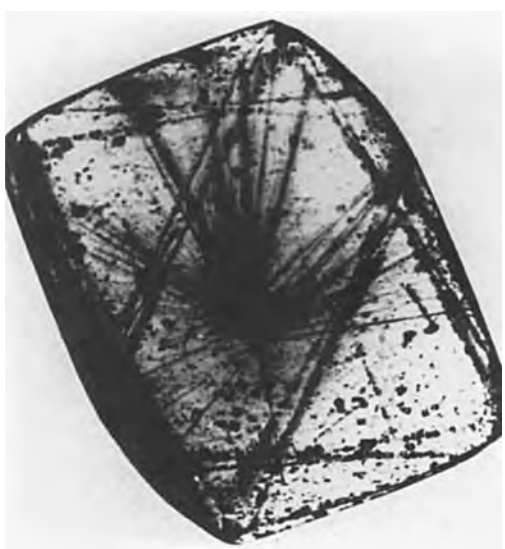


Figure 17. X-ray topograph of a diamond containing dislocations radiating in all directions from an internal flaw (Lang [63]).

Figure 18. X-ray topographs illustrating the perfection of certain diamonds (Lang [63]).

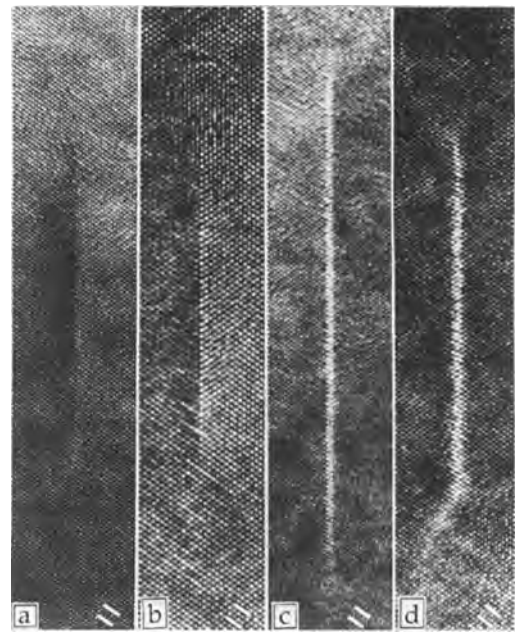


Figure 20. (a) to (d) High resolution images of nitrogen platelets in diamond (HREM) (courtesy of W. Luyten).

are responsible for the yellow color of 'cape' diamonds. High resolution images (Fig. 20) have been used to determine the atomic structure [72, 73] by comparing simulated images of various proposed models with the observed images.

Some natural diamonds contain tiny octahedral-shaped cavities called voidites [74–80]. In TEM they sometimes produce a strain pattern, proving that the voidite causes stress in the surrounding diamond matrix. Moiré patterns (Fig. 21a) give information on the lattice parameters of the material enclosed in the void relative to the lattice parameter of diamond. A peculiar form of double diffraction allows us to deduce indirectly the single crystal diffraction pattern of the material in the void [81], which is found to be a form of solid nitrogen.

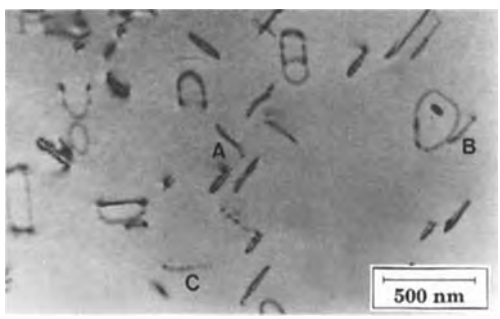


Figure 19. Diffraction contrast image (TEM) revealing nitrogen platelets in diamond. The platelets are situated in cubic planes (courtesy of W. Luyten).

Concentric growth layers can be observed in many natural diamonds by means of cathodoluminescence using a scanning electron microscope (see, e.g., Goldstein and Yakowitz [82]). These patterns are due to slight variations in the impurity contents of the material out of which the diamond has grown (blue ground). Since no two diamonds have identical growth patterns, the

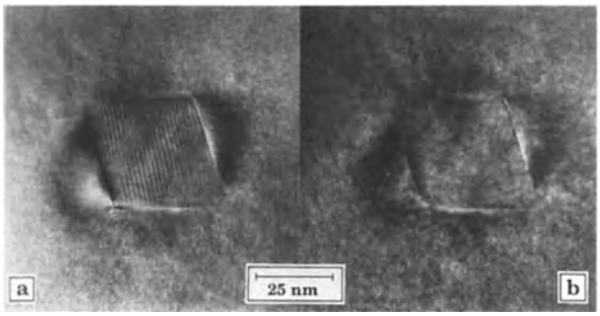


Figure 21. (a), (b) Voidites in natural diamond. Note the presence of Moiré fringes in (a). The strain field in the surrounding diamond matrix is imaged as well (Luyten et al. [81]).

cathodoluminescence pattern can be used as a 'fingerprint' for a given stone in forensic applications. Since dislocation lines or centers located close to them act as recombination centers, they emit light in cathodoluminescence (Fig. 22). For a general review of defects in diamonds, we refer to Wilks and Wilks [3].

11.4.2 Synthetic Diamond

A synthetic diamond single crystal grown by means of the high pressure, high temperature method, is shown in Fig. 23, as observed in SEM.

In recent years, with the discovery of plasma enhanced, low pressure, chemical vapor deposition methods to grow epitaxial layers of diamond, the way was opened for the application of diamond films in solid state electronics and in the wear protection of tools [3–7]. This has triggered a number of studies on the degree of perfection of such layers using high resolution electron microscopy (HREM) (e.g., references [83–97]). The films are often prepared by epitaxial growth on a (1 1 1) silicon single crystal base, in view of applications in microelectronics. Owing to the lattice misfit with the substrate, the diamond layers are often heavily faulted. Specimens consisting of a free,

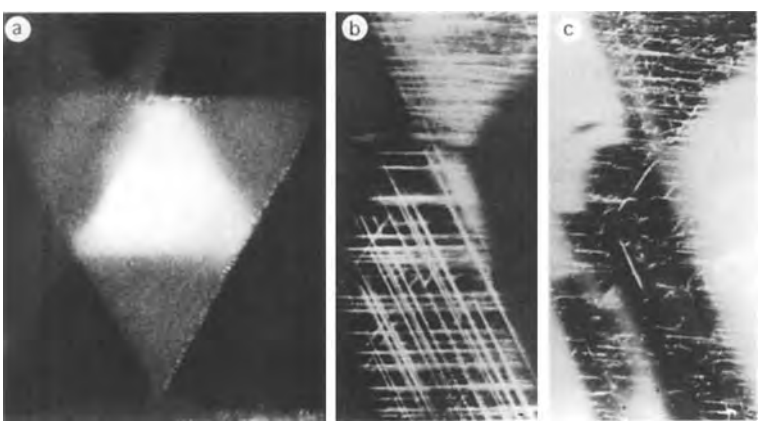


Figure 22. Cathodoluminescence in natural diamond. (a) Imaging the impurity distributions. (b) Area containing dislocations. (c) Same area as (b), imaged by X-ray topography.

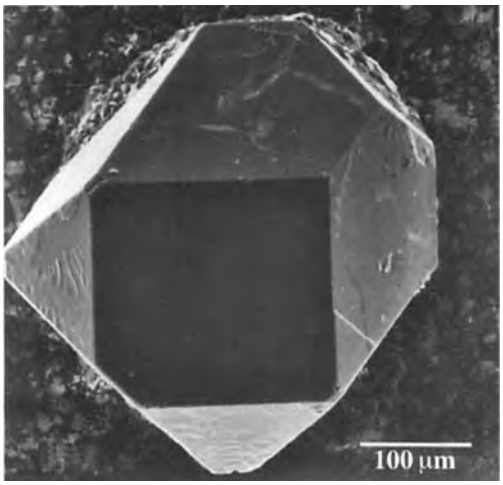


Figure 23. Synthetic diamond prepared by the high pressure, high temperature method (SEM).

unsupported diamond film of uniform thickness parallel to a (1 1 1) face can easily be made by selectively dissolving the silicon base. Such studies revealed the occurrence of various defects related to the relatively small stacking fault energy of diamond.

The most prominent defects are twin boundaries (Fig. 24). Frequently sequences of the order of ten parallel interfaces, separating relatively narrow twin lamellae, can be observed edge-on. Attempts to measure the lattice relaxation along such twin interfaces were inconclusive [97]. Steps in twin boundaries, forming twinning dislocations, are often found as well. Twin boundaries, inclined with respect to the foil surface, separate overlapping, twin-related crystal parts (Fig. 25). In the overlap region a coincidence dot pattern is observed in the HREM image, exhibiting a triple period compared to that of diamond; owing to double diffraction, the diffraction pattern also exhibits the same spacing tripling (Fig. 25) [97, 98].

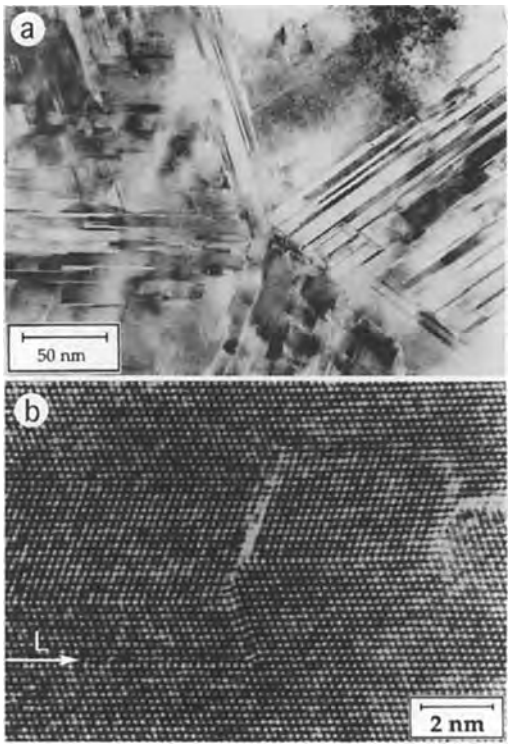


Figure 24. Low magnification, high resolution image along the [1 1 0] zone of a free synthetic diamond flake prepared by plasma-enhanced CVD on a silicon substrate. Note the high concentration of twins and stacking faults on the (1 1 1) planes (HREM). (a) Five twin-related grains forming a disclination along their common line. (b) Twins and a Lomer–Cottrell barrier in L (Luyten et al. [97]).

Glide dislocations on the (1 1 1) planes are dissociated in Shockley partials connected by a strip of intrinsic stacking fault. Such a configuration (Fig. 26) can be used to deduce the stacking fault energy. The value obtained in this way was $\gamma = 209 \pm 26\,\mathrm{mJ\,m^{-2}}$, to be compared with that deduced from natural crystals by the weak beam method ($\gamma = 279 \pm 41\,\mathrm{mJ\,m^{-2}}$ [65]). This discrepancy was attributed to surface effects in the weak beam case, where the dislocations are parallel to the surfaces (see Sec. 11.3.2 of this Chapter).

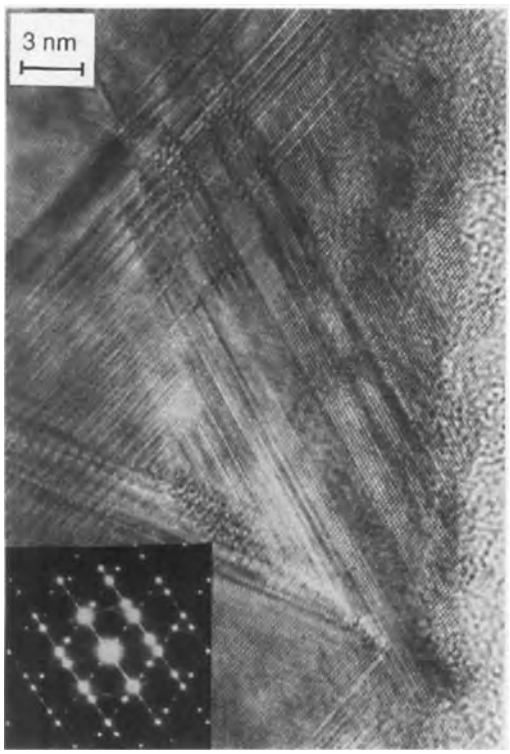


Figure 25. Overlapping twin lamellae in synthetic diamond simulating a triple period in the area of the overlap (HREM). Inset is the corresponding diffraction pattern along the [1 1 0] zone, also exhibiting a threefold period (courtesy of L. C. Nistor).

11.5 Fullerenes

11.5.1 C_{60} Crystal Structures: Phase Transitions

The room temperature structure of solid C_{60} of sufficient purity is a face-centered arrangement (cubic, $a_0 = 1.41$ nm) of quasi-spherical molecules with radius $r = 0.353$ nm (see, for example, Heiney et al. [17]). The apparent sphericity results from the (almost) free rotation of the molecules at room temperature.

It was found by electron diffraction and microscopy that crystals grown from a solution were faulted hexagonally closed-packed, but transform in situ into a faulted face-centered cubic structure, presumably as a consequence of evaporation, in the microscope vacuum, of solvent molecules (Fig. 27) [11].

As a result of the particular ratio of the lattice parameter and the radius of the 'spherical molecules', the $h\,0\,0$, $0\,k\,0$, and $0\,0\,l$ reflections are accidentally extinct even for h, k, and l: even, as was demonstrated by means of X-ray diffraction [99] and electron diffraction [100] (Fig. 28). This may lead to a sensitive method to study the phenomena that cause a change of this ratio and thus lead to the appearance of these reflections, for h: even (k: even or l: even) [101].

Figure 29 shows an HREM image of a perfect C_{60} crystal seen along the [110] zone. At 255 K an orientational phase transition occurs; the structure remains cubic but the symmetry decreases; the highest symmetry space group $Fm\bar{3}m$ is reduced to the subgroup $Pa\bar{3}$, the three fourfold axes being replaced by twofold axes [16, 17, 102]. The symmetry transla-

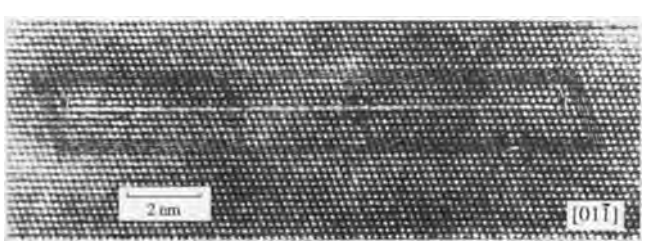


Figure 26. Glide ribbon consisting of Shockley partials separated by an intrinsic stacking fault (Burgers circuits are outlined). Such ribbons are used as probes to measure the stacking fault energy (HREM) (Luyten et al. [97]).

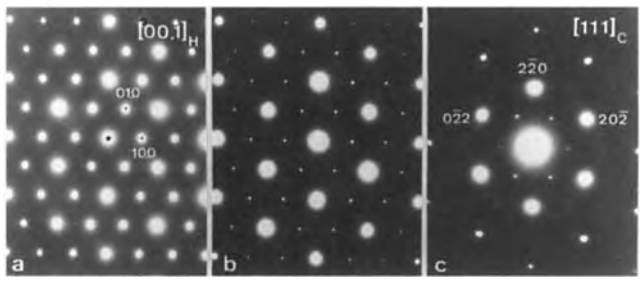


Figure 27. Evolution at room temperature in the microscope of the diffraction pattern of a crystallite of C_{60} grown from an organic solvent. The diffraction patterns all refer to the same area of the crystallite after increasing time intervals in the microscope. (a) Predominantly hexagonally packed. (b) The relative f.c.c. stacking frequency increases. (c) Predominantly f.c.c. stacked with faults (Van Tendeloo et al. [11]).

tions of the type $1/2\langle 110\rangle$ are lost since the lattice becomes primitive with approximately the same lattice parameter. A single crystal is hereby fragmented in two orientation variants related by a 90° rotation about a cubic axis. Moreover, each orientation can occur in four translation variants, having the lost symmetry translations as displacement vectors. The electron diffraction patterns of Fig. 30 prove the presence of two orientation variants [14, 103]. In particular they illustrate the loss of fourfold symmetry of a single variant; the symmetry breaking superstructure spots are somewhat weaker than the basic spots. In Fig. 30c two orientation variants are present in the selected area. The corresponding orientation domains can be imaged by making dark field images in superlattice reflections. The domain walls are found to coincide often with (111) twin interfaces [14, 103].

At still lower temperatures, certain crystals were found to exhibit a further superstructure with a lattice parameter $2a_0$ [104]; however, this was not confirmed by X-ray diffraction [105].

X-ray and neutron diffraction have shown that the $Pa\bar{3}$ superstructure is due to the 'freezing in' of the molecules in four different orientations, each of the four primitive sublattices of the face-centered cubic lattice being occupied by molecules

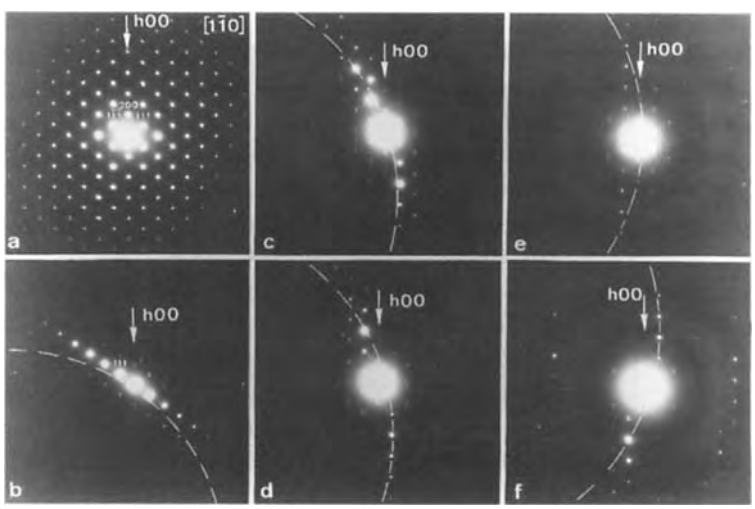


Figure 28. Room temperature diffraction experiment on C_{60} showing the extinction of the $h00$ diffraction spots with h: even. (a) to (f) On gradually tilting, Ewald's sphere is swept through the row of $h00$ nodes (indicated by an arrow), without exciting 200, 400, etc., although such reflections are allowed in the f.c.c. lattice. In (a) such reflections are produced by double diffraction (Amelinckx et al. [100]).

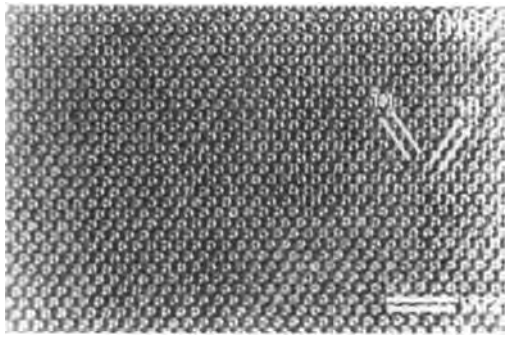


Figure 29. High resolution image of a perfect C_{60} crystallite as viewed along [1 1 0] (Muto et al. [107]).

having the same orientation [17]. The relative orientations of adjacent molecules are determined by a simple packing principle: parts of the cage surface where there is an accumulation of electric charge face parts of the neighboring cage where there is a deficiency of electric charge, thus minimizing the total coulomb energy.

11.5.2 Defects in F.C.C. Crystals of C_{60}

Crystals of the room temperature form of C_{60} exhibit the same kinds of defects as are typically found in f.c.c. alloys with a very

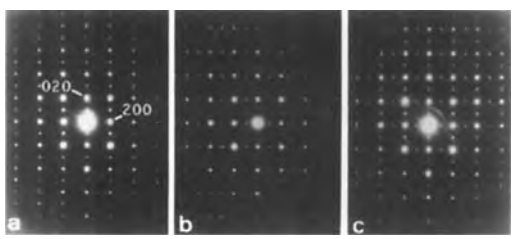


Figure 30. [0 0 1] zone diffraction pattern of the simple cubic low temperature phase of C_{60}; note the twofold rotation symmetry of the pattern due to the presence of weaker superstructure spots. In (a) one variant is revealed, in (b) the second orientation variant is selected, and in (c) both variants are present in the selected area (Van Tendeloo et al. [14]).

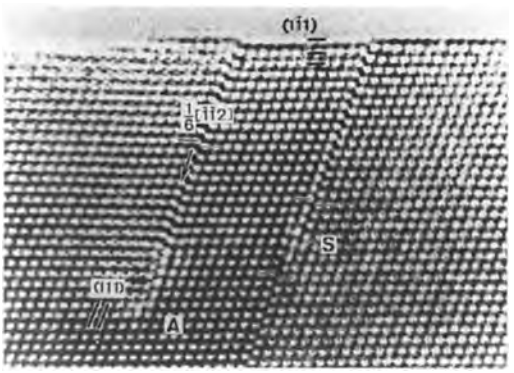


Figure 31. Stacking fault (S) and Frank dislocation and its associated fault (A) in C_{60}; zone axis [1 1 0] (Muto et al. [107]).

small stacking fault energy, the molecules playing the role of the atoms in the alloys [13, 14, 103, 106, 107]. Growth phenomena on the surface of single crystals were observed by optical microscopy [108].

The dominant defects revealed by HREM are coherent twin interfaces and stacking faults on (111), both intrinsic and extrinsic, and their combined configurations, with partial dislocations such as Lomer–Cottrell locks, and Shockley and Frank dislocations (Fig. 31) [107]. Owing to the exceptionally small stacking fault energy, defects resulting from the intersection of intrinsic stacking faults can also be formed, as shown in Fig. 32: these defects are linear arrays of 'fractional' vacancies.

11.5.3 C_{70} Crystal Structures: Phase Transitions

At room temperature, vapor grown crystals of very pure C_{70} (>99.9%) exhibit, even within the same batch, two different morphologies [109]. The majority of the crystals is cubic, but a minority has a

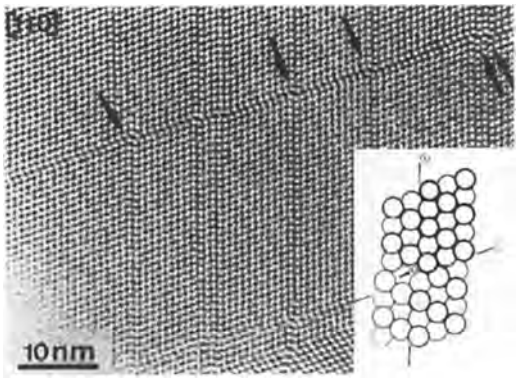


Figure 32. Intersecting stacking faults in C_{60} (zone axis [1 1 0]). Inset: model for the intersections (Van Tendeloo et al. [103]).

hexagonal habit. The corresponding room temperature structures, as deduced from electron diffraction and HREM, are face-centered cubic and hexagonally close-packed, both structures being heavily faulted respectively on $(1\,1\,1)_{f.c.c.}$ or $(0\,0\,0\,1)_{h.c.p.}$ planes (Fig. 33) [103, 110].

As first shown by means of in situ low temperature electron diffraction, the two types of crystal undergo orientational phase transitions on cooling. At liquid nitrogen temperature, the c/a ratio of the h.c.p. form is $c/a = 1.82$, which corresponds to the axial ratio for an h.c.p.

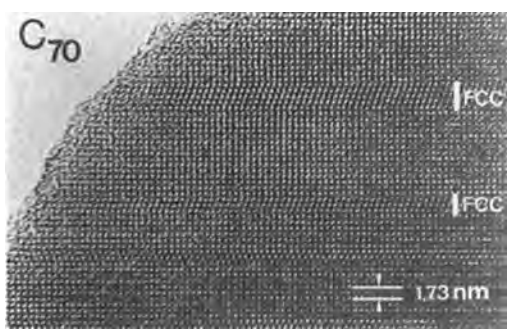


Figure 33. Room temperature HREM image along $[1\,1\,0]_{f.c.c.}$ of a crystallite of C_{70} exhibiting a mixture of hexagonal and cubic stacking (Muto et al. [107]).

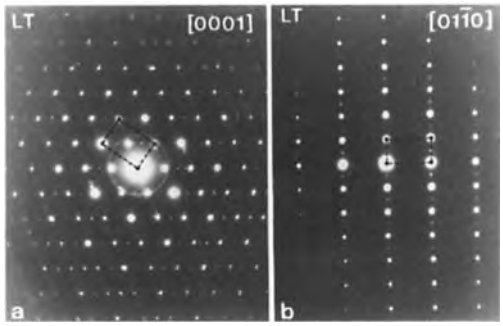


Figure 34. Electron diffraction patterns of the low temperature phase of the hexagonal form of C_{70}. (a) [0 0 0 1] zone and (b) [0 1 $\bar{1}$ 0]. Note the parameter doubling (Van Tendeloo et al. [110]).

stacking of ellipsoids of revolution with their long axis parallel to c and with an aspect ratio corresponding to that of C_{70}. Moreover, the a-parameter of the h.c.p. basic lattice is doubled along one direction, breaking the hexagonal symmetry (Fig. 34a). The lattice can be described either as monoclinic with $\beta = 120°$ or as C-face-centered orthorhombic. It was shown that in this low temperature phase the ellipsoidal C_{70} molecules are h.c.p. with their long axis parallel to c and orientationally ordered within the layers according to the same principle as mentioned above for the low temperature phase of C_{60} [103, 109, 110]. As the temperature rises above 276 K, the superstructure spots disappear, but the c/a ratio remains roughly the same. This was interpreted as resulting from the rotation of the molecules about their long axis, which still remains parallel to c. At around 337 K, the c/a ratio decreases rapidly and approaches a value close to that for the packing of spheres ($c/a = 1.63$). This was taken to mean that the molecules now freely rotate, hereby simulating a spherical shape. Two different stacking modes of the same

ordered layers, one of them leading to doubling of the *c*-parameter (Fig. 34b), were observed in the low temperature phase, suggesting the possibility of orientational polytypism [110].

The f.c.c. stacked form of C_{70} has a low temperature form which is slightly rhombohedral, with a temperature-dependent rhombohedral angle, the long axis of the ellipsoids being oriented along the long diagonal of the elongated rhombohedron. The transformation temperature is slightly above room temperature. By varying the electron beam intensity, the crystal structure can be switched in situ from rhombohedral to f.c.c. and vice versa [103, 107]. The f.c.c.–rhombohedral transition gives rise to four orientation variants corresponding to the four possible orientations of the ⟨1 1 1⟩ direction, which becomes the long diagonal of the rhombohedral phase. A single crystal of f.c.c.-stacked C_{70} thus breaks up in a twinned domain structure on cooling; this was shown by optical microscopy (Fig. 35a) and on a finer scale by AFM, as shown in Fig. 35b [23].

These results are in good agreement with those of other techniques, such as X-ray diffraction [111–117], neutron diffraction [118], and NMR (nuclear magnetic resonance) observations [119, 120]. However, considerable undercooling of the different phases and the existence of several metastable phases (such as the h.c.p. phase at room temperature) complicated the interpretation of the experiments considerably [115, 121].

11.5.4 Intercalated C_{60}

Intercalates of C_{60}, such as $C_{60}(I_2)_{2-x}$ [122] and Rb_6C_{60} [123], were imaged by means of HREM. The iodine enters the C_{60} as a neutral molecule. The in situ deintercalation of $C_{60}(I_2)_{2-x}$ by the loss of I_2 molecules in the microscope vacuum was studied in detail and shown to proceed by the migration of dislocation-like defects bordering the intercalated areas, causing a shear over $\frac{1}{2}c$ along a prismatic plane of the simple hexagonal structure, followed by relaxation along a direction normal to this prismatic plane. In the wake of the moving dislocation, the AA stacking of the simple hexagonal structure of the C_{60} matrix of the intercalate is transformed into the ABC... stacking of the f.c.c. structure of pristine C_{60} (Fig. 36) [122]. This mechanism explains the orientation relationship between the intercalated and the pristine phases.

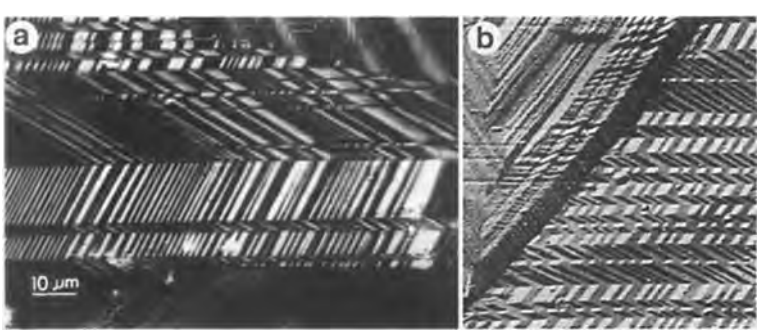


Figure 35. Domain structure in rhombohedral C_{70} crystallites as revealed by means of (a) optical microscopy and (b) AFM (Verheyen et al. [23]).

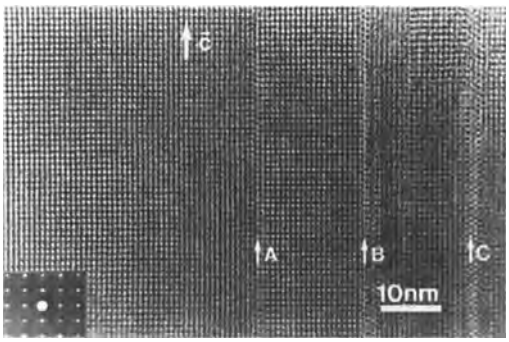


Figure 36. Prismatic faults and bands A, B, C of partly disintercalated simple hexagonal $C_{60}(I_2)_{2-x}$ in HREM. Inset: corresponding diffraction pattern; note the streaking (Zhang et al. [122]).

In Rb_6C_{60} and in similar alkali-doped C_{60} intercalates, the alkali atoms are ionized and form well-defined superstructures. These compounds are very sensitive to the humidity of the air. Handling the sample in view of its preparation as a microscope specimen and of its subsequent insertion in the microscope requires special precautions. An inexpensive, but nevertheless efficient method, making it possible to keep the material under a protective gas atmosphere during all these operations, was described in Zhang et al. [123]. HREM images of the Rb_6C_{60} intercalated at two different defocus conditions are reproduced in Fig. 37 and compared with the structure model.

Atomic resolution STM studies of K_xC_{60} and Ca_xC_{60} thin film intercalates show the dependence of the surface structure on the potassium or calcium concentration [19, 124]. C_{60} molecules are revealed as featureless 'domes' in STM images of pure C_{60} films at room temperature, due to the rapid rotation of the molecules. However, images obtained from Ca_xC_{60} thin films clearly show intramolecular features for some values of x, reflecting the cage structure of each individual molecule, and thus indicating that the molecules have a fixed orientation

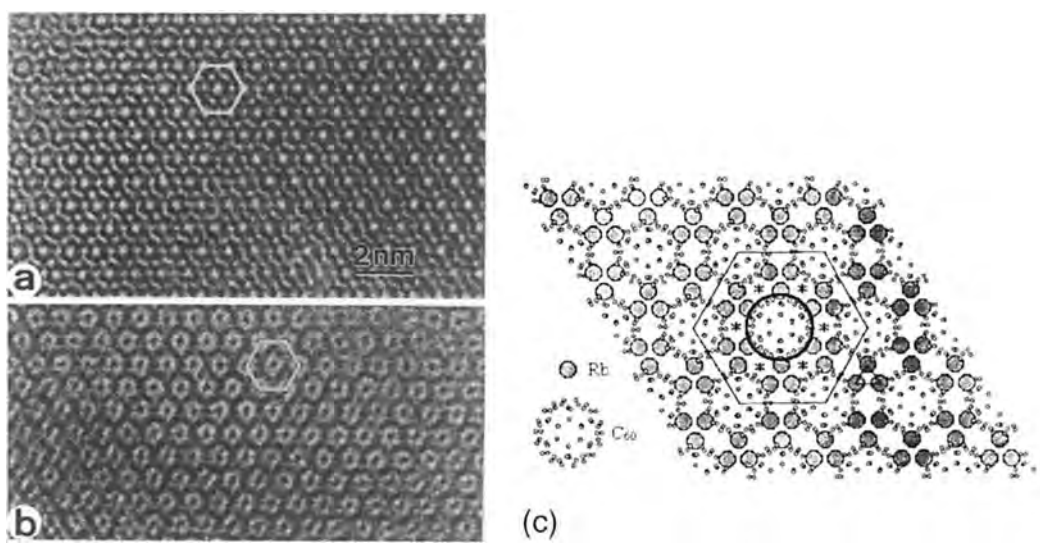


Figure 37. HREM along [1 1 1] of rubidium intercalated C_{60}, compared with (c) a schematic diagram of the structure. Images (a) and (b) were taken at two different defocus values (Zhang et al. [123]).

(similar observations are reported for C_{60} films at 4.5 K [125]). The positions of the potassium ions could not be imaged directly, since the STM images were found to show the changes in the electronic structure due to the potassium intercalation, rather than the real structure of the film surfaces.

11.5.5 C_{60} and C_{70} Thin Films

Epitaxial thin films of C_{60} and C_{70} have been prepared by vapor deposition on cleaved surfaces of many substrates, such as metallic, semiconducting, and layered semiconducting materials. Atomic resolution AFM and STM are well suited for the surface structure determination of the first monolayers, revealing the growth mode, structure, defects, and orientational degrees of freedom of the fullerene molecules [18–24, 126, 127]. TEM and electron diffraction can be used successfully to study the structure and quality of somewhat thicker films [128–134] and to obtain the orientation of the overlayer on the substrate surface [135, 136]; no observations of cross-sectional samples have been reported to date. At present, the best quality films have been obtained on one of several lamellar substrates such as GeS (0 0 1), GaSe (0 0 0 1), MoS_2 (0 0 0 1), and mica (0 0 1).

Germanium sulphide (GeS) is a particularly interesting substrate for the deposition of C_{60} films. It cleaves easily along the (0 0 1) plane and then exhibits a surface that is grooved along one direction [137]. The distance between the grooves happens to be precisely half the equilibrium separation of close-packed rows of C_{60} molecules, making it possible to obtain good epitaxial layers by systematically 'filling' every other groove. The orientation relationship is then $[1\,1\,1]_{C60} // [0\,0\,1]_{GeS}$ and $[1\,0\,\bar{1}]_{C60} // [0\bar{1}0]_{GeS}$ [136–138]. The geometry of the epitaxy may induce 'antiphase boundaries', either by missing out two grooves instead of one, or by a lateral shift of the sets of occupied grooves (Fig. 38). These features were found by means of TEM, which is a well-established method of obtaining the orientation relationship from a composite foil: substrate plus film. Antiphase boundaries were revealed in HREM [136].

11.6 Carbon Nanotubules and Nanoparticles

11.6.1 Introduction

Carbon nanotubules generally consist of a concentric packing of hollow, cage-like carbon configurations. Carbon nanoparticles are spherical or polyhedral shaped [139], whereas carbon nanotubes are cylindrical structures, closed by caps and with lengths several orders of magnitude larger than their diameter [27]. The structure of the caps (Fig. 39b) can be considered as graphene sheets, in which the removal or the insertion of 60° wedges, generating disclinations in the graphene sheet, causes curvature of the otherwise flat layers. The cylindrical part, on the other hand, consists of elastically deformed graphene sheets (Fig. 39a).

Experimental techniques such as X-ray diffraction, which probe the average structure of a macroscopic quantity of material,

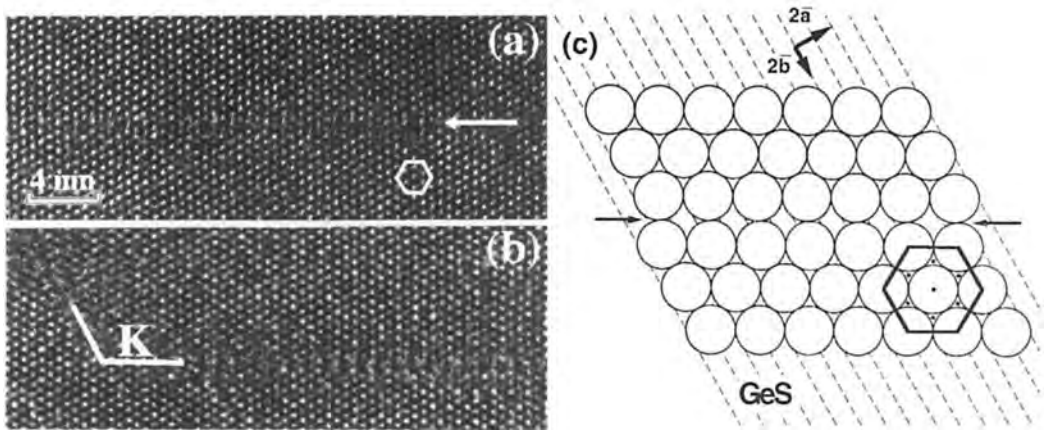


Figure 38. (a) and (b) Out-of-phase boundaries in epitaxial films of C_{60} on a (0 0 1) cleavage face of GeS (the zone axis is $[1\,1\,1]_{C60}$; the lattice of GeS is not resolved) (HREM). The hexagon connects the centers of C_{60} molecules of the first monolayer, as also indicated in (c). Note that only one third of the bright dots correspond to the centers of molecules of the first monolayer (A-type layer). The other dots represent the centers of the molecules of the B-type and C-type monolayers. (c) Schematic representation of the out-of-phase boundary (Bernaerts et al. [136]).

are not well suited for the study of such carbon materials. SEM is able to show the morphology of the material [140–143]. AFM [144, 145] and STM [144, 146, 147] were used to study the external shape and the atomic structure on the surface of the tubes. However, most structural information about carbon nanotubes results from the application of TEM techniques, such as high resolution imaging and electron diffraction, although TEM is not able to show the individual carbon positions due to the absence of well-defined atomic columns [27, 30–32, 148–159].

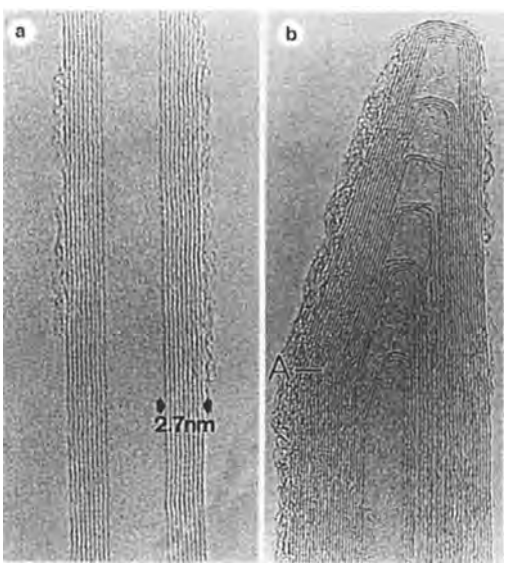


Figure 39. Multishell carbon nanotubes prepared by the arc discharge method and imaged by means of {0 0 0 2} lattice fringes ($d_{0002} = 0.344$ nm). (a) Straight tube. (b) Terminal caps (Zhang et al. [153]).

Until recently, there were two major routes for the production of carbon nanotubes: nucleation from a carbon plasma in an electric arc [27, 141, 143, 160, 161] or the catalytic decomposition of hydrocarbons [39, 162, 163], each having several variants. The catalytic method is very similar to the production of mesoscopic carbon fibers, already known for a long time (see, for example, references [164–168]). Very recently, variants of these methods leading to single-walled tubes have become available [36, 37]. First we will discuss observations of perfect structures and finally observations of structural defects

in these carbon configurations will be shown.

The hollow channel inside the carbon nanotubules can be filled with various materials, as was observed by TEM [169–176]. The filling takes place either during production of the tubules, or by opening the tip of the tubes by a chemical reaction and subsequently filling them with liquid phase, due to capillary forces. Owing to the small space available inside the channel, the crystal structure of the filling material can be different from that of the normal bulk material.

11.6.2 The Microstructure of 'Perfect' Carbon Particles

The structure of a single shell carbon cage is completely determined by the positions of the pentagons in the otherwise hexagonal graphene layer [177, 178]. A carbon nanotube consists of a cylindrical 'body', exclusively containing hexagons, terminated by two caps, each containing six pentagons. The aromatic structure of the cylindrical body is completely determined by only three parameters: the length L and the radius r, giving the dimensions of the tube, and the chiral angle θ, representing the orientation of the graphene network with respect to the tube axis [27] (negative chiral angles are allowed, taking into account the hand of chiral tubes). L and r can easily be obtained from HREM images; the chiral angle θ can be determined from the electron diffraction patterns. The sign of θ and thus the hand of the tubes can be obtained from diffraction contrast images observed under a suitable incidence direction of the electron beam. For more theoretical purposes, it may be suitable to replace r and θ by the pair of integers, the Hamada indices (n, m) (see Appendix C).

Determination of the Radius r and Length L: High Resolution Electron Microscopy (HREM) Images

Generally, images of carbon nanotubes are recorded with the tube axis normal to the electron beam (normal incidence), showing the two walls and the hollow central channel (Fig. 39). The main features are the {0002} lattice fringes imaging the edge-on view of the graphene layers in the tube walls [27]. The fringe spacing is 0.344 nm, compatible with the (0002) interplanar distance in turbostratic graphite [179]. Since most tubes are relatively small (<10 nm), and since carbon is a weak electron scatterer, carbon nanotubes can be considered as thin, weak phase objects. At Scherzer defocus, HREM images of such samples show dark contrast at sites with a high projected electron density. The dark fringes can thus be directly associated with graphene layers seen edge on. The number of layers in a tubule and the radii of the tubes can thus be readily determined. Image simulations confirm that the dark {0002} fringes correspond to the graphene planes (for Scherzer defocus) [153, 180, 181], but measuring the distance between the fringes in an image of a single shell tube is not a very accurate method for the determination of r, as was shown in Wang and Zhou [182].

HREM images with two-dimensional resolution sometimes show $\{1\,0\,\bar{1}\,0\}$ fringes (spacing 0.21 nm) in the central and wall region (Fig. 40). The orientation of this

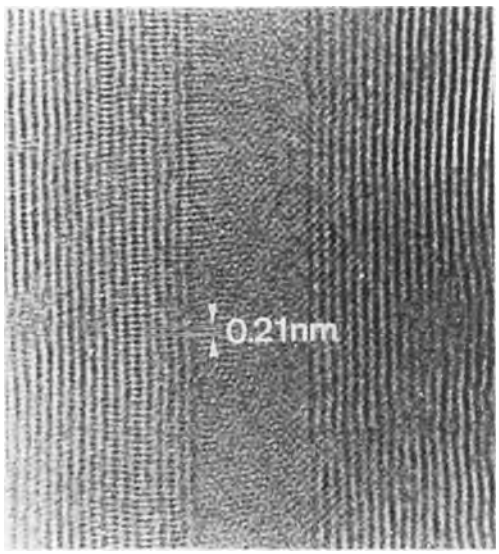


Figure 40. High resolution image of a multishell tubule exhibiting, next to the {0 0 0 2} fringes, $\{1\,0\,\bar{1}\,0\}$ fringes ($d_{10\bar{1}0} = 0.21$ nm) (Zhang et al. [153]).

pattern with respect to the tube axis can be used to determine the chiral angle [153]. By examining the contrast and symmetry of the *hk*.0 fringes, it is possible to draw some conclusions about the cross section of the tubules [183, 184]: for a perfectly cylindrical tubule, the angle between the electron beam and the graphene layers changes continuously as a function of the projected distance to the tubule axis. Consequently, the appearance of the $\{1\,0\,\bar{1}\,0\}$ fringes in the HREM image is expected to change continuously as a function of the distance to the tubule axis. For tubules with a polygonized cross section, however, large parts of the tubule have the same orientation with respect to the electron beam, independent of the distance to the tubule axis, and the fringe contrast is expected to be constant along strips parallel to the tubule walls. As shown by Fan and Bursil [181], it is in principle possible to determine the chiralities of every single shell tube in a multishell tubule from HREM images exhibiting $\{1\,0\,\bar{1}\,0\}$ fringes, as long as the number of layers is small. The power spectrum of the digitized image is obtained by fast Fourier transformation, clearly revealing the different chiralities in a multishell tubule. By allowing only the spots corresponding to a specific chirality to contribute to the reconstructed image (obtained by inverse Fourier transformation), the chirality can be assigned to a specific layer of the multishell tuble. However, the high resolution required, problems with radiation damage, deviations from normal incidence, and defects in the carbon structure severely limit the practical use of such images for the structure determination of carbon nanotubes.

The outer radius and length of nanotubes can also be determined from AFM or STM images of tubes firmly attached to a substrate such as silicon [145–147]. Owing to the dependence of the images on the shape of the AFM or STM tip, the most accurate results are obtained from height measurements, rather than from the lateral dimensions of the tube images. Atomic resolution images allow the determination of the helicity, but only of the outermost tube.

Determination of the Chiral Angle θ: Electron Diffraction (ED) Patterns

The chirality of carbon nanotubes can be derived from HREM, AFM, or STM images, although with many practical difficulties and limitations. High resolution Moiré patterns, due to overlap of the top and bottom parts within the projected area of the channel, can provide an alternative means of determining the chiral angle

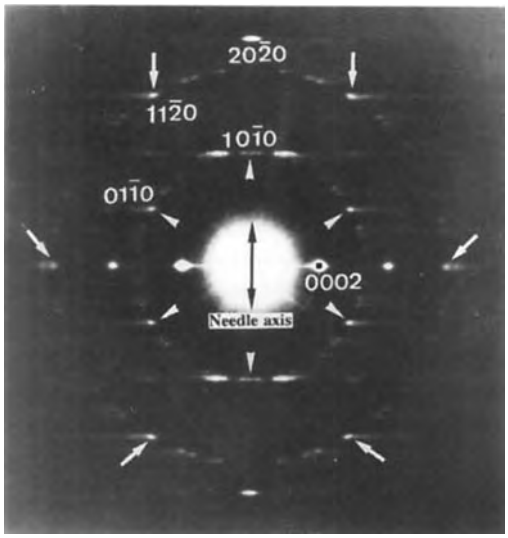


Figure 41. Electron diffraction pattern of a multishell nanotubule made under quasi-normal incidence. Note the weak spots due to chiral tubes; most of the tubes are achiral, however, as deduced from the relatively higher intensity of the spots indicated by arrows (Zhang et al. [155]).

[184]. However, these HREM techniques require normal incidence of the electron beam, a condition that cannot be recognized from the images alone. Electron diffraction (ED) patterns obtained from an isolated tubule (Fig. 41) give both the angle of incidence of the electron beam and the chiralities present in the tubule [see Eqs. (3) and (4)]. The tubule that produced Fig. 41 was largely achiral, as deduced from the main spots (arrowed), but weaker spots with the same spatial frequency are split, proving the presence of chiral components.

The interpretation of ED patterns developed for carbon nanotubules can also be applied to that of other cylindrical structures such as chrysotile serpentines [185] or cylinders of MTS_3 misfit layer compounds [186]. (M = Pb, Sn, ...; T = Nb, Ta, ...)

The helicity θ of a carbon nanotube can conventionally be defined as the smallest angle enclosed by the C–C bonds and the needle axis [sometimes it is defined as $\pi/(2-\theta)$]. In achiral tubes θ is zero for zigzag or parallel tubes and 30° for armchair or normal tubes. As a result of the hexagonal symmetry of the graphene sheets, other C–C bonds enclose an angle $(\theta + 60°)$ or $(\theta - 60°)$ with the tubule axis.

A simple intuitive interpretation of ED patterns valid for normal beam incidence was given by Iijima [27]. The stacking of the graphene layers in the wall regions is represented as a row of equidistant $0\,0\,0\,l$ (l:even) reflections. The hexagonal lattices of the layers in the top and bottom part of the tubule causes hexagonal sets of $hk.0$ spots in the diffraction pattern. For nonhelical tubes, the patterns of the top and bottom parts superpose and only one hexagonal pattern is visible. For helical tubes, the hexagonal $hk.0$ pattern is rotated with respect to the $(0\,0\,0\,l)$ row; the pattern corresponding to the top part is rotated in one sense, while the pattern of the bottom part is rotated in the other sense. Consequently, two hexagonal sets of spots are visible in the ED pattern, their angular separation being a measure for the chirality θ of the tube. Owing to the cylindrical nature of the tubes, the $hk.0$ spots are clearly streaked outwards, away from the tubule axis, as explained in more detail below and in Appendix D (Fig. 41). Diffraction patterns observed for single shell carbon nanotubes show the same features, only the $0\,0\,0\,2$ spots are absent [33, 187].

This intuitive interpretation is only valid for normal incidence, and cannot explain the behavior of the spots upon tilting the tube. Moreover, using patterns

not obtained under normal incidence, the method for measuring the chirality will give inaccurate results if not corrected for inclination [188]. Therefore consideration of the complete three-dimensional diffraction space is necessary [155].

Construction of diffraction space: It is possible to imagine several structural models of multiwall carbon nanotubes, either 'scroll' type or concentric cylinder type, for the construction of diffraction space. We call 'ideal' a scroll in which the layers forming the successive turns are superposed without relaxation of the atomic structure of the layers, in particular the interlayer spacing being rigorously conserved. The increase in length of successive turns ($2\pi \times$ layer thickness per turn) is then accommodated by a continuously increasing relative shear of successive curved, but otherwise undeformed layers. In a 'real' scroll this shear is not uniform since atoms in one layer tend to occupy 'hollows' in the underlying layer, leading to strips of good fit separated by partial edge dislocations of the type presented in Fig. 6 parallel to the fiber axis [153]. If these dislocations are arranged in radial planar walls, the scroll acquires a polygonal cross section. If, on the contrary, the dislocations are roughly randomly distributed (with the same average spacing) along the turns, the scroll's cross section remains to a good approximation a circular Archimedian spiral. In the latter case it is justified to consider the stacking of the tangential c layers as disordered by random shifts parallel to the layers, the shift vectors being determined by the Burgers vectors of the partial dislocations. We discuss the geometry of diffraction space for these two extreme cases. The case of achiral tubules, consisting of concentric tubes, can be modeled, as far as the diffraction effects are concerned, as real scrolls. Chiral tubules are considered separately.

1. Ideal scroll—We first discuss the diffraction space of ideal scrolls using a model similar to that used in the literature [189]. Figure 42a shows two successive

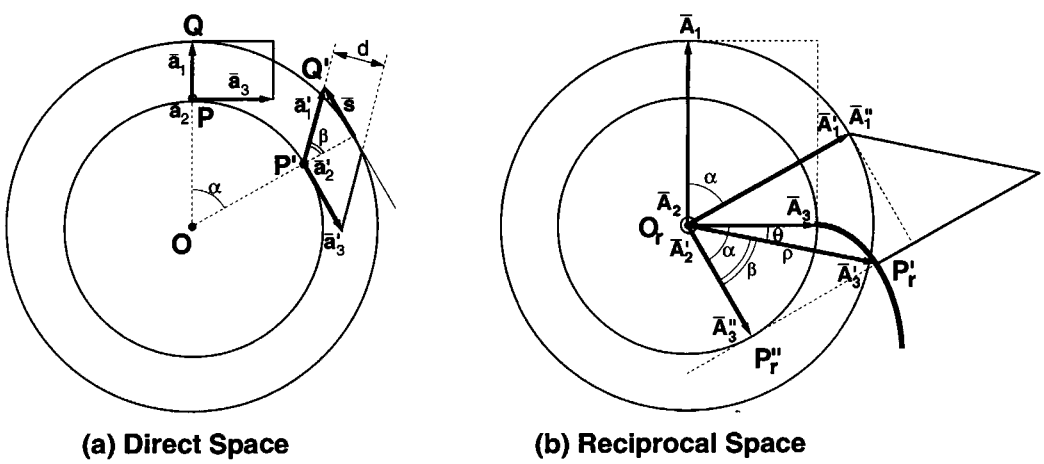


Figure 42. Illustration of the notations used for the derivation of diffraction space for an ideal scroll: (a) direct space; (b) reciprocal space.

layers in the cross section of an achiral circular cylindrical tubule. Let the unit cell of the structure at P be constructed on the base vectors $(\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3)$. On rolling over an angle α in such a way that $\overline{PP'} = \overline{QQ'}$, the relative shear is $\boldsymbol{s}$ for two successive layers. The unit mesh $(\boldsymbol{a}_1, \boldsymbol{a}_3)$ is sheared over an angle β. The new base vectors are $(\boldsymbol{a}'_1, \boldsymbol{a}'_2, \boldsymbol{a}'_3)$. The vector $\boldsymbol{a}_2 = \boldsymbol{a}'_2$; $\boldsymbol{a}'_1$ encloses an angle β with the radial direction, $\boldsymbol{a}'_3$ is tangential to the inner circle and is thus parallel to the direction of shearing.

Since the tube is assumed to be achiral, our geometrical considerations can be limited to a plane normal to the fiber axis; the problem is then reduced to two dimensions. Let the reciprocal unit mesh of $(\boldsymbol{a}_1, \boldsymbol{a}_3)$ be $(\boldsymbol{A}_1, \boldsymbol{A}_3)$ (Fig. 42b). After rolling over an angle α this unit mesh occupies the orientation $(\boldsymbol{a}'_1, \boldsymbol{a}'_3)$ in direct space and $(\boldsymbol{A}''_1, \boldsymbol{A}''_3)$ in reciprocal space. However, we have hereby as yet ignored the shear $\boldsymbol{s}$ over an angle β. Taking this into account the unit mesh in reciprocal space becomes $(\boldsymbol{A}'_2, \boldsymbol{A}'_3)$, where $\boldsymbol{A}'_3$ is perpendicular to the plane $(\boldsymbol{a}'_1, \boldsymbol{a}'_2)$ and has a length equal to the universe of the direct space interplanar distance between these planes. From Fig. 42b $|\boldsymbol{A}'_3| = 1/d$; $d = |\boldsymbol{a}_3| \cos\beta$ and hence $|\boldsymbol{A}'_3| = |\boldsymbol{A}_3|/\cos\beta$.

It is our objective to find the geometrical locus of the reciprocal lattice node P'_r at the endpoint of $\boldsymbol{A}'_3$. We note that $\overline{P'_r P''_r} = |\boldsymbol{A}_3| \tan\beta$ and that the arc $\overline{A''_3 A_3} = |\boldsymbol{A}_3|\alpha$. Since, moreover, $|\boldsymbol{s}| = 2\pi|\boldsymbol{a}_1|(\alpha/2\pi) \equiv |\boldsymbol{a}_1|\alpha$ and, moreover, $\tan\beta = |\boldsymbol{s}|/|\boldsymbol{a}_1|$ (from Fig. 42a), $\tan\beta = \alpha$ and hence $|\overline{P'_r P''_r}| = |\overline{P''_r A_3}|$, from which we conclude that the geometrical locus of P'_r is described by a point on the straight line $P'_r P''_r$ when the latter rolls without sliding on the circle with radius $O_r A_3$. Such a curve is called the evolute of a circle. The reciprocal lattice nodes corresponding to families of lattice planes parallel to the axis of the fiber thus describe circle evolutes.

In the system of polar coordinates (ρ, θ) with origin in O_r, the locus of the point P'_r is obtained as follows, $\theta = \alpha - \beta$ and $\rho = |\boldsymbol{A}'_3| = |\boldsymbol{A}_3|/\cos\beta$ (from Fig. 42b); since $\tan\beta = \alpha$ and $\cos\beta = (1 + \alpha^2)^{-1/2}$, we finally have $\theta = \alpha - \arctan\alpha$ and $\rho = |\boldsymbol{A}_3|\sqrt{1 + \alpha^2}$, which is the equation of a circle evolute in polar coordinates with α as a parameter. If α is large enough, approximately $\theta = \alpha - \pi/2$ and $\rho = A_3\alpha$ or $\rho = A_3\theta + const$, which represents an Archimedean spiral with a separation A_3 between successive turns.

2. Real scrolls and tubules—In real scrolls and even more so in achiral tubules consisting of concentric tubes, the structures in successive turns (tubes) are no longer always shifted systematically over the same displacement vector as was the ideal case for homogeneous shear, but such shifts, parallel to the outer planes, are now random. As a result, the scroll (or tubule) can now be assumed to consist of volume elements of one-dimensionally disordered c-layers (as in graphite), which are always at a tangent to the scroll wall (see Fig. 43). The diffraction space of one such volume element then consists of a two-dimensional array of *hk*.0 nodes of the layer structure in a plane parallel to the fiber axis. Except for the origin, all such nodes are streaked along a direction normal to the (0001) layer plane, i.e., along c^*. Only the $000l$ row consists of sharp nodes since $000l$ reflections are not affected by shifts parallel to the $(000l)$ planes; these nodes describe circles with radii lc^*.

If the one-dimensional disorder is not completely random, reinforcements will

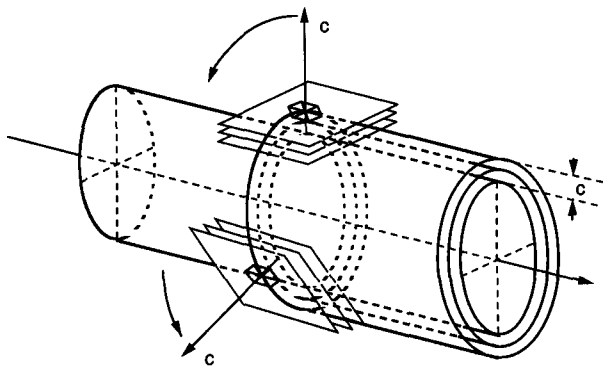


Figure 43. Model used in deriving diffraction space for a real tubule (or scroll) (Zhang et al. [155]).

occur in the streaks; these correspond to preferred periodicities in the arrangement of layers. In monolayer tubules the streaking of the *hk*.0 nodes, as well as that through the origin, is due to the 'shape' effect, i.e., to the extreme 'thinness' of the layer or, stated otherwise, to the two-dimensional nature of the tangent 'volume' elements.

The reciprocal space of the scroll (or tubule) now consists of the geometrical loci traced out by the reciprocal space of one such block on rotation about an axis through the origin parallel to the fiber axis. The streaks hereby generate 'disks' or 'coronae' of diffuse scattering sharply limited inwards by the circles described by the *hk*.0 nodes, and fading outwards. A schematic view of the reciprocal space for an achiral scroll (or tubule) is shown in Fig. 44. An intuitive direct space model for diffuse intensity is discussed in Appendix D.

As an ideal scroll or a multilayer tubule is assumed to become more real, the set of evolutes evolves into the coronea of diffuse

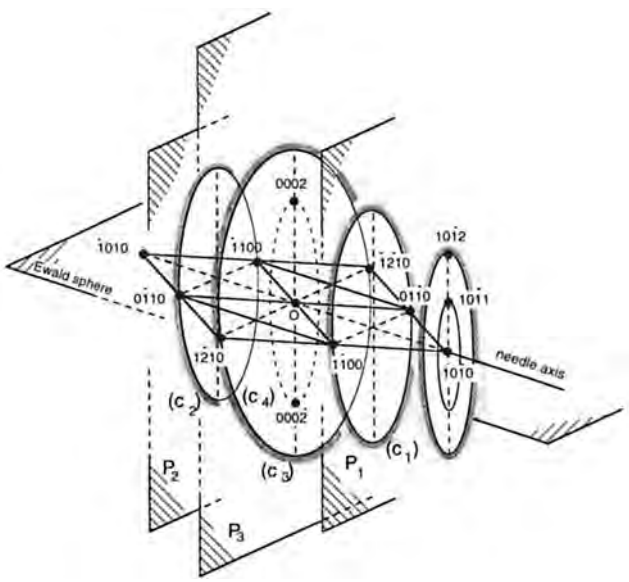


Figure 44. Schematic representation of diffraction space for an achiral tubule. Each node describes a circle in diffraction space (Zhang et al. [155]).

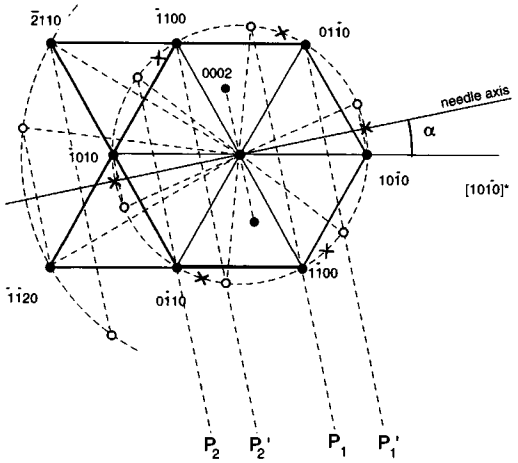


Figure 45. Schematic representation of spot doubling (full and open circles) characteristic of the diffraction space for a chiral tubule (Zhang et al. [155]).

scattering. The diffraction patterns of multilayer carbon fibers can satisfactorily be interpreted on a real model. In achiral tubes the situation is degenerate in the sense that two (or more) nodes give rise to a single 'coronae'. In chiral tubes this degeneracy is lifted and each node describes its own circle (and coronea) obtained by rotation about an axis parallel to the needle axis of Fig. 45.

The diffraction pattern for an arbitrary orientation of the tubule is found by considering the intersection of Ewald's sphere (plane) with the diffraction space. Both the tilt angle α and the true helicity θ can be found from a single diffraction pattern by means of the following relations [188]

$$\tan\theta = \frac{\cos\phi_1 - \cos\phi_2}{\tan\gamma(\cos\phi_1 + \cos\phi_2)} = -\frac{\tan\left(\frac{\phi_1 - \phi_2}{2}\right)\tan\left(\frac{\phi_1 + \phi_2}{2}\right)}{\tan\gamma} \qquad (3)$$

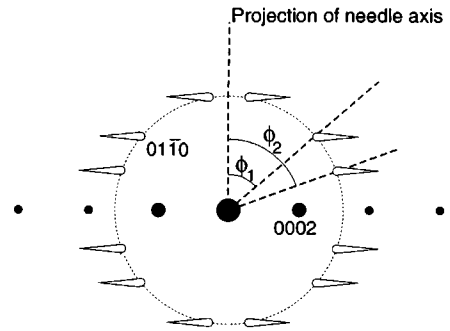


Figure 46. Illustration of the symbols used in Eqs. (3) and (4).

and

$$\cos\alpha = \frac{\cos(\gamma - \theta)}{\cos\phi_1} \equiv \frac{\cos(\gamma + \theta)}{\cos\phi_2} \qquad (4)$$

with ϕ_1 and ϕ_2 the measured angles in the inclined position between the spots of the doublet and the projected tube axis (Fig. 46); γ is the angle for the considered spots in the case of an achiral tube at normal incidence. At normal incidence $\gamma = 0$, 30°, 60°, or 90°, depending on whether spots of the first or second hexagon are considered.

The Hand of Chiral Tubes: Diffraction Contrast Dark Field Images

ED patterns are insensitive to the hand of chiral tubes, as are the HREM images of {0002} fringes. However, for specific incidence angles of the electron beam (for example, tilt-angle $\alpha = \theta$), Ewald's plane is tangential to one of the *hk*.0 circles in diffraction space leading to the coincidence of two reflections. Dark field diffraction contrast images taken in the coinciding *hk*.0 reflection, which is located on the line perpendicular to the row of *c*-reflections (see Fig. 47a), allows the determination of the hand of chiral nanotubes (it

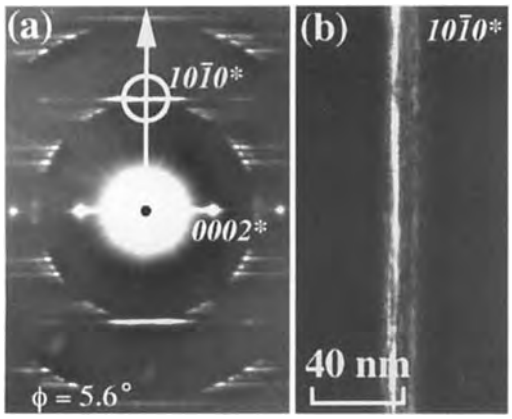


Figure 47. (b) Tilted dark field image used to determine the hand of chiral tubules and (a) the corresponding diffraction pattern (Bernaerts et al. [189a]).

must be stated that a clear definition of the hand of a tube is required: by convention, we will define the hand of a chiral tube as that of the helical string of hexagons through a given point of the cylindrical surface which encloses the largest angle with the tubule axis, i.e., which has the smallest pitch) [189, 190]. For specific tilting angles, and by selecting the appropriate diffraction spot (as indicated on Fig. 47a), the brightness of both walls in the dark field diffraction contrast image differs; which wall is the brightest depends on the sense of tilting and on the chirality of the tube (Fig. 47b). This dark field technique can be used for multishell, multichiral tubes, and is able to provide a map of the distribution of the chiralities inside a multishell tubule, although with relatively poor resolution [189a].

A similar technique was used in the past to determine the hand of chiral tubular textures such as tobacco mosaic virus [191], MoS_2 tubes [192], and the hand of screw dislocations [193].

11.6.3 Carbon Onions

Intense electron irradiation of the carbon clusters results in considerable changes in the carbon network and can be a problem for high resolution observations. However, experiments in which the carbon nanotubes are intentionally irradiated were performed by Ugarte [194–196]. After very intense irradiation, the carbon nanotubes transformed during the first stage from cylindrical structures into amorphous material consisting of graphene tube fragments. Longer irradiation transformed this material into spherical particles, which are concentric stackings of spherical cages with systematically increasing diameter: so-called carbon onions (Fig. 48a). The innermost cage is of the same size as fullerene globular molecules such as C_{60}. From careful comparison between simulated and experimental images, it was concluded that such carbon onions are spherical in contrast to cage-like single shell C_n globular particles, which would be polyhedral [180]. The Van der Waals interaction between different layers in a carbon onion favors the spherical shape above the polyhedral shape [198]. The relative orientation of the single shell molecules forming the onion was found to be random [180].

A quite remarkable phenomenon was recently discovered: carbon onions can apparently be used as nanoscale pressure cells, capable of sustaining pressures (~50–100 GPa) and temperatures high enough to transform graphite into diamond under equilibrium conditions [199a]. Carbon particles such as nanotubes, when electron irradiated in situ (at temperatures of the order of 750°C)

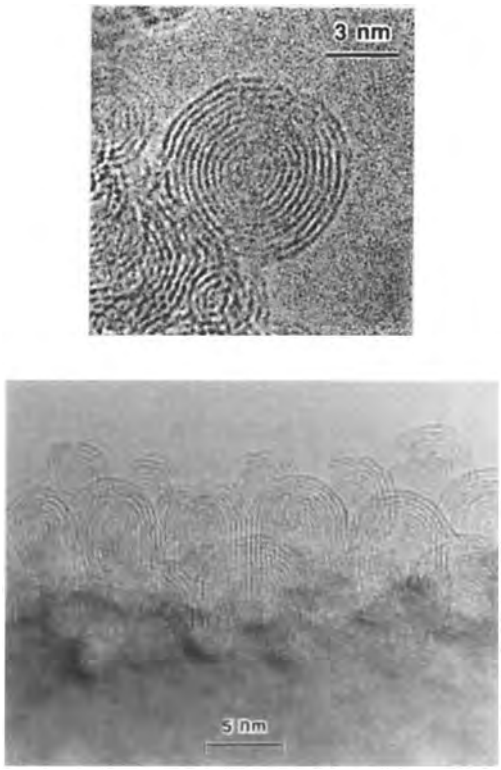


Figure 48. (a) Carbon onions containing several concentric shells (Ugarte and de Heere [197]), (b) Onions formed under intense electron irradiation at the surface of diamond. There is a directional correlation of the epitaxial onion-like graphitic particles and the diamond substrate. The (00.1) plane of graphite is well aligned with one family of (111) planes of diamond [199b]. (Courtesy of S. Iijima)

in a high voltage (1250 kV) microscope, transform into large onions, containing numerous (>15) concentric spherical shells. The increased atom mobility due to the combined effect of electron irradiation and high temperature allows the radiation damage to anneal out as it is formed. The pressure in the central part of the onion can be deduced from the systematic decrease in c-spacing (from 0.32 nm down to 0.22 nm) towards the center, as observed by HREM. On prolonged irradiation at temperatures between 650 and 750°C, the central cavity becomes filled with a nanocrystal of diamond, identified by high resolution imaging and electron diffraction. Irradiation at room temperature produces radiation damage, but no diamond.

Recently [199b] the transformation of sp_2 bonded graphitic structure was studied 'in-situ' using the electron microscope simultaneously as a heat source, as a small accelerator and as an imaging device. A thin chip of diamond was irradiated at 200 keV at an electron dose rate of about 10^7 el/nm^2 sec ($\sim$ 150/cm^2). After a total dose of about 10^9–10^{10} el/nm^2 high resolution images of the thin edge of the specimen exhibit 'onions' containing 4–10 graphitic shells, superposed on the diamond lattice image (Fig. 48b).

This nanostructure appears to be due to the dynamic equilibrium between on the one hand the production of atom displacements resulting from electron excitations, rather than from the direct impact of electrons, and on the other hand the thermal energy and the charging effect favouring atomic migration and restructuring along the surface. It should be noted that the carbon displacement threshold in diamond is 80 eV corresponding to an electron energy of 330 keV, which is above the operating electron energy of the microscope. It is further noted that these graphite particles are not stable under less intense electron irradiation, such as used in high resolution electron microscopy (of the order of 10^6 el/nm^2 sec). This is a consequence of the smaller carbon displacement threshold in graphite which is only about 30 eV (corresponding to $\sim$ 140 keV).

11.6.4 Defects In and Deformation of Carbon Nanotubules

Radiation Effects

Recently, more attention is being given to the study of defects in carbon particles. During HREM observations, defects are induced by the intense electron radiation. Therefore useful observations can only be made for a short time and with relatively low electron energies. Recently, it was found that heating the sample to 300°C during observation, limits the effect of radiation, even for electron energies of 1.25 MeV, due to the higher defect mobility in the carbon network [200].

Deformation of Carbon Nanotubes

Elastic bending and the resulting deformation of the tubes under the influence of external forces were recently studied by HREM [154, 201, 202]. The inner rim of the tubules, which is under compressive stress, was found to exhibit a number of kinks, depending on the bending angle. Although the graphite layers are locally severely deformed, the elastic limit is not exceeded, as was proven by a TEM experiment which showed complete restoration of the original shape, in agreement with molecular dynamics simulations.

The radial deformation of carbon nanotubes was revealed by careful analysis of 0002 fringe images of juxtaposed tube pairs [203]. The slightly deviating interfringe spacing in the contact region was attributed to elastic deformation of the two nanotubes due to the Van der Waals attraction between them.

'Fully collapsed' carbon nanotubes have been identified by TEM in soot prepared by the carbon arc method [204]. The cross section of these multilayer tubes is not circular but consists of a flattened ribbon like part where the graphite layers are parallel, terminated on both sides by strongly curved borders. Such a shape would result when a circular multishell tubule is completely flattened; it is stabilized by the Van der Waals attraction between the parallel flattened parts. The shape was deduced from the presence of twists and from in situ rotation experiments around the length axis.

Very recently [205], low magnification transmission electron microscopy has been used to estimate Young's modulus of single multishell nanotubes. It is found that tubes supported from one end only, as in a cantilever, vibrate constantly, independently of the intensity of the electron beam. This phenomenon is attributed to the thermal energy associated with the ambient temperature. Focussing of the vibrating end is then impossible. However, the extreme positions, i.e., the vibration amplitude, can be recorded as a function of temperature in a heating holder. From this temperature dependence, Young's modulus can be estimated. It is found to be in the TPa range, which is much larger than that of carbon fibers used hitherto for the reinforcement of composite materials.

Pentagons and Heptagons

Pentagonal meshes in the hexagonal graphene network are necessary to close the carbon cage (Fig. 39). Also, the presence of heptagonal meshes has to be inferred from the observed negative curvature in certain parts of the graphene layers. The effect of pentagons and heptagons on

the external shape of a tube was observed by TEM [148, 149, 152, 157, 206, 207] and AFM [145]. A pair consisting of a heptagon and a pentagon in diametrically opposed positions causes a tube to be bent over an angle of approximately 30°, and can induce a change in the radius and helicity [189, 208–210].

A pentagon–heptagon pair in adjacent positions in a graphene sheet forms a pair of 60° disclinations of opposite sign. Such a pair of disclinations is equivalent to an edge dislocation having its supplementary 'half-line' parallel to the line connecting the centers of the two disclinations. Such defects can therefore account for changes in the diameter of straight cylindrical tubules; they would topologically be necessary in seamless conical tubes having semi-apex cone angles different from those given by Eq. (9). Shape and size changes of tubules, as observed by means of {0 0 0 2} fringes, have been explained by assuming combinations of pentagons and heptagons [149].

Helically Coiled Tubules

The periodic insertion of heptagon–pentagon pairs in diametrically opposed sites on the tubule can transform a straight carbon nanotube into a torus or into a helically coiled tubule, which were theoretically shown to be energetically stable structures [211]. Such structures are believed to be observed in the material deposited during the thermal decomposition of hydrocarbons, using finely dispersed metal particles supported by amorphous silica as a catalyst [38, 39]. TEM observations revealed that the deposited material contains a large number of multishell carbon nanotubes, a significant fraction being regularly helically coiled (Fig. 4). The radius R of the helix ranges from 8 nm to a few hundred nanometers; the pitch p of the helix ranges from 20 nm to a few hundred nanometers. A method to derive these parameters for a helix in an arbitrary orientation is described in Bernaerts et al. [159]. The handedness of the helix can be derived by observing the helical tube with the incident electron beam inclined with respect to the helix axis. The difference in shape of the helix parts on both sides of the helix axis and knowledge of the tilt direction enables us to determine the handedness of the helix; a method that is more or less similar to one of the methods to determine the handedness of chiral tubes.

A model for the diffraction space of helically coiled nanotubes is summarized in Appendix E. In the diffraction pattern, the 0 0 0 l reflections become arcs; their angular extent depending on the shape of the helix and on the tilt angle of the helix axis with respect to the electron beam. This was confirmed by ED and TEM images, and the model is in agreement with diffraction contrast dark field images such as shown in Fig. 49 [159]. For a perfectly helically coiled tubule, a continuous intensity distribution is expected in the 0 0 0 l arcs. However, experimental ED patterns obtained from a single turn of the helix clearly show that the 0 0 0 l arcs have a spotty appearance if a single turn of the helix is selected (Fig. 50d), indicating that the helices are polygonized. This was confirmed by TEM images [159]. The bending angle was found to be approximately 30°, the same value as expected for a tube with a heptagon–pentagon pair inserted in diametrically opposed sites in the graphene network of the tube

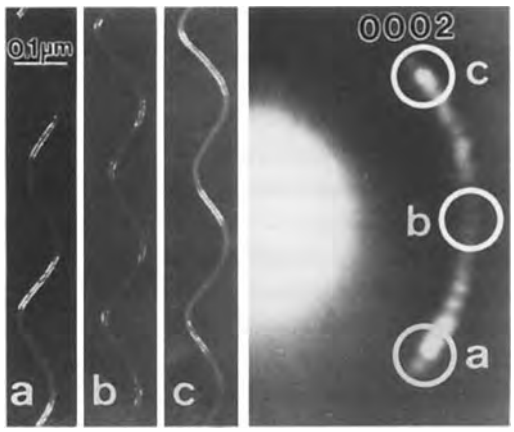


Figure 49. Diffraction contrast dark field images (a), (b), and (c) made in different parts of the 0002 diffraction arcs of a helical tubule, as indicated in the diffraction pattern. The two walls are revealed as bright segments (Bernaerts et al. [159]).

[205, 206]. Based on morphological observations and on the observations of catalytic particles in association with the tubules, a growth model accounting for the helical shape could be proposed [212].

Cross Section of Tubules

Most HREM images of carbon nanotubules show the same number of fringes in both walls. Therefore Iijima concluded that the tubules must be concentric stackings of closed cylindrical tubes: the 'Russian doll' model [27]. However, this model often seems to be oversimplified for two reasons: (i) the cross section of a considerable fraction of the multishell tubules deviates from circular, and (ii) there are indications that some tubes are not a concentric arrangement of closed tubes, but are, at least in part, scrolls. This was concluded from the presence of anomalous interfringe spacings [32] (Fig. 51).

As already mentioned, it is impossible for a given graphene layer in a carbon nanotube to remain in register with the structure of the previous layer due to the circumference difference $2\pi(c/2) = \pi c$, even if both tubes have the same chirality.

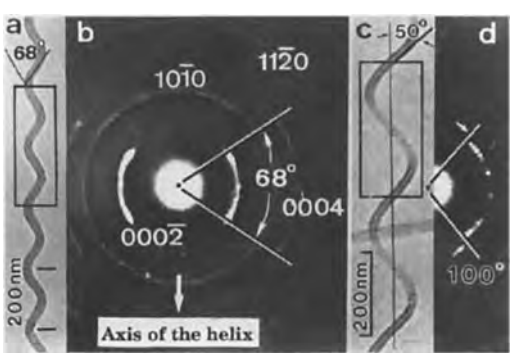


Figure 50. Electron diffraction pattern of a polygonized helical tubule. In (a) and (b) several turns are included in the selected area, whereas in (c) and (d) only one turn is included (Bernaerts et al. [159]).

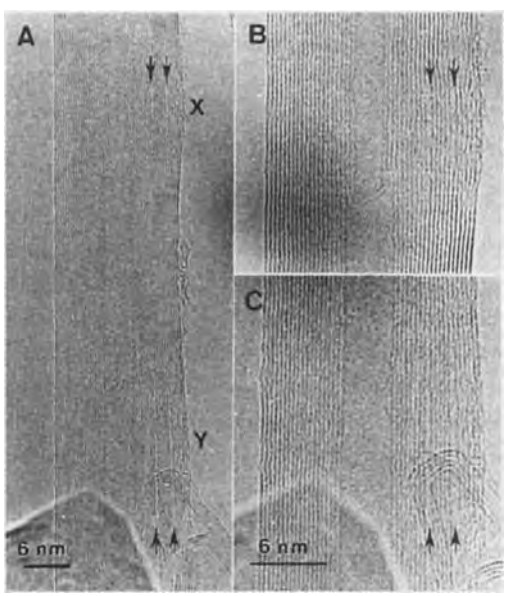


Figure 51. Anomalously spaced lattice fringes, indicated by arrows, in a multishell tubule attributed to dislocations terminating scroll parts in the tubule. (B) and (C) are magnified parts of X and Y in (A) (Amelinckx et al. [32]).

The cross section of a tubule can be circular or polygonized, as stated above. However, from the two-dimensional projected images obtained by TEM, it is not straightforward to distinguish between both cases, if the electron beam is not parallel to the tubule axis. From images taken at normal incidence, in combination with extensive tilting about the tube axis, it is clear that a considerable part of the tubules has a cross section deviating from a circular one, especially close to the cap of the tube (see, for example, Li et al. [213]). Liu and Cowley [156, 158] concluded from nanodiffraction patterns that some nanotubes have a polygonized cross section, and interpreted HREM images of tubes exhibiting some extra wide, singular 0002 interfringe spacings in one wall, such as in Fig. 51, as being an indication of polygonized tubes (this feature was reinterpreted more recently, however [32]). Furthermore, AFM and STM observations confirmed that tubes with circular, as well as with polygonal cross sections exist [144, 145]. Recently, TEM images were obtained from tubes viewed along their axis by embedding the tube containing soot in an epoxy resin and by subsequently sectioning the samples with ultra microtomy (UMT) [154, 214, 215]. Although it is difficult to distinguish between cross sections of cylindrical nanotubes and more or less spherical nanoparticles, these images are in agreement with the AFM and previous TEM results.

By recording ED patterns and HREM images of the same tubule, the number of different chiral angles observable in the ED patterns are found mostly to be much smaller than the number of layers in the tubule (observations of monochiral tubes of more than 10 layers were reported), suggesting that such tubules may be scrolls. Furthermore, intercalation experiments revealed that metal atoms can penetrate and destroy the layer structure of the tubes [31]. These two results are unlikely for the Russian Doll model for carbon nanotubes, and an alternative model was proposed: carbon nanotubes can consist, at least partly, of one or more graphene sheets rolled into scrolls, thus leading to a cross section similar to an Archimedian spiral [32]. (This model was also proposed in the past for carbon fibers [164].) All turns of a scroll have of course the same chirality. A terminating or starting scroll introduces an edge dislocation in the tubule wall, and it was argued that the deviating graphene layer distance in the close neighborhood of these edge dislocations causes the singular 0002 fringe spacings observed in some HREM images [32] (Fig. 51). For the configurations of the edge dislocations to be as expected on theoretical grounds, the number of fringes must remain equal in both walls. Furthermore, terminating 0001 layers were observed in HREM images taken for normal incidence [145] and in cross-sectional images [154, 214–216].

11.6.5 Conically Shaped Carbon Configurations

Cylindrical carbon nanotubes are observed to be closed by dome-shaped caps or by conical caps depending on the configuration of the pentagons in the graphene structure (Fig. 39b). STM [217], SEM, and TEM [40, 41] observations

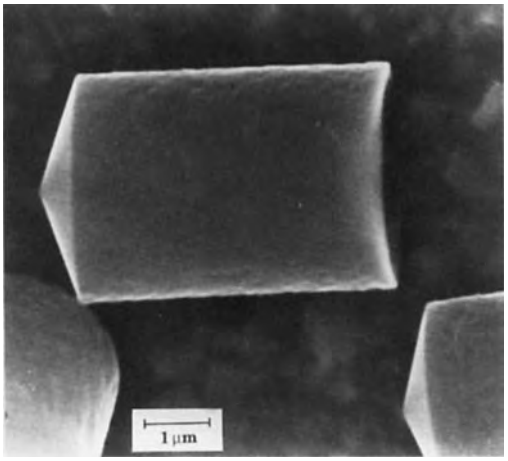


Figure 52. SEM image of a conically wound carbon needle formed by the decomposition of CO on an SiC substrate. The fracture surfaces show the conical character of the needle (TEM). The semi-apex angle is ~70° (Amelinckx et al. [40]).

showed the existence of isolated, conically shaped carbon structures (Fig. 52), and 'bamboo shaped' carbon nanotubes consisting of a string of conical structures were observed by HREM [218]. As shown in Appendix F and in the literature [40, 41, 217], only certain semi-apex angles are allowed for seamless cones. Different semi-apex angles correspond to a different number of pentagons in the tip of the cone. Cone angles slightly deviating from these specific values were seen to occur in conical graphite whiskers [40, 41]. Since for these observed semi-apex angles no seamless cones are expected, the whisker was found to consist of a conical graphene plane, helically wound around a 'disclination' with a fivefold carbon ring core. Each turn of a conical scroll is rotated with respect to the previous turn over an angle $\theta = 2\pi \sin \phi_0$ (ϕ_0 being the semi-apex angle of the cone) about the local c-direction. This rotation of each successive layer with respect to the previous one explains the peculiar diffraction effects observed for such conical whiskers, as shown in Luyten et al. [41].

The diffraction pattern of conically wound single fibers, the incidence beam being parallel to the cone axis, is much more complicated than that due to cylindrically wound fibers [219]. The most prominent feature is the presence of two rows of equidistant sharp $000l^*$ reflections through the origin, enclosing an angle $2\phi_0$. These rows are perpendicular to the projected wall directions. They are, moreover, both shifted along their length towards the origin, as a result of 'refraction' of the electron beam at the walls of the cone. The other reciprocal lattice node points generate loci which are spherical spirals, i.e., curves situated on spheres of constant spatial frequency.

In general each node generates a spherical spiral, but in cases of high symmetry and for special values of ϕ_0, all nodes of a given spatial frequency may generate a single spiral. For a detailed discussion, we refer to Amelinckx et al. [219].

11.7 Appendices

11.7.1 Appendix A: Euler's Rule

C_{60} and C_{70} are convex polyhedra which obey Euler's topological rule, i.e.,

$$V + F - E = 2 \tag{5}$$

where V is the number of vertices, E the number of edges, F the number of faces, h the number of hexagons, and p the number of pentagons in one polyhedron. For this particular case, it can be derived very

simply. Obviously for the number of edges

$$5p + 6h = 2E \tag{6}$$

with each pentagon contributing five edges and each hexagon six edges; however, since each edge belongs to two faces we have counted the number of edges twice. The number of faces is obviously

$$F = p + h \tag{7}$$

Since each edge connects two vertices and along each vertex three edges come together,

$$E = (3/2)V \tag{8}$$

because again we have counted each edge twice. Eliminating p and h between (6), (7), and (8), keeping in mind that $p = 12$ (see App. B), leads to (5). This relation is quite generally valid for convex polyhedra, also if more than three edges come together in a vertex.

11.7.2 Appendix B: Curvature Introduced by Pentagonal Meshes

The introduction of a pentagonal mesh in a hexagonal honeycomb network can be modeled using a continuum model as the extraction of a 60° wedge followed by closing of the resulting gap, i.e., as the introduction of a negative 60° disclination. The planar sheet is hereby deformed into a cone with a semi-apex angle ϕ_0 related to the extracted wedge angle θ (θ, e.g., 60°) by the relation [40]

$$\theta = 2\pi \sin\phi_0 \tag{9}$$

This cone covers a solid angle Ω_0 in units 4π

$$\Omega_0 = \frac{1}{2}\int_0^{\phi_0} \sin\phi \, \mathrm{d}\phi = \frac{1}{2}(1 - \cos\phi_0) \tag{10}$$

Of more direct interest is the cone formed by the local normals to this first cone. The corresponding solid angle Ω_n is given by the same equation [Eq. (10)], but the integration limits are different

$$\Omega_n = \frac{1}{2}\int_{\pi/2}^{\pi/2-\phi_0} \sin\phi \, \mathrm{d}\phi$$

$$= \frac{1}{2}\sin\phi_0 = \frac{\theta}{4\pi} \tag{11}$$

For $\theta = 60°$, $\Omega_n = 1/12$, i.e., twelve pentagons are necessary to form a closed cage fullerene. The simplest polyhedron is the pentagon dodecahedron which contains exclusively 12 pentagons.

11.7.3 Appendix C: Relation Between r, θ, and (n, m) for Chiral Tubes

The atomic structure of the cylindrical part of a seamless single shell carbon nanotube is determined by the radius r and the chiral angle θ. Consider a line along the complete circumference of the tube and perpendicular to the tube axis; let the starting point A and the endpoint A′ of the line coincide (Fig. 53). By cutting the tube along a line parallel to the tube axis and through A (and A′), the tube can be unfolded to a flat, two-dimensional graphene sheet. The segment AA′ is then defined as the chiral vector $\boldsymbol{c}_\mathrm{h}$; it connects two crystallographically equivalent sites on the graphene sheet. This chiral vector $\boldsymbol{c}_\mathrm{h}$ completely characterizes the geometry of the tube [178], and by specifying $\boldsymbol{c}_\mathrm{h}$ in terms of the unit lattice vectors of the honeycomb lattice ($\boldsymbol{a}_1$ and $\boldsymbol{a}_2$), the integer Hamada indices (n, m) are obtained by the relation

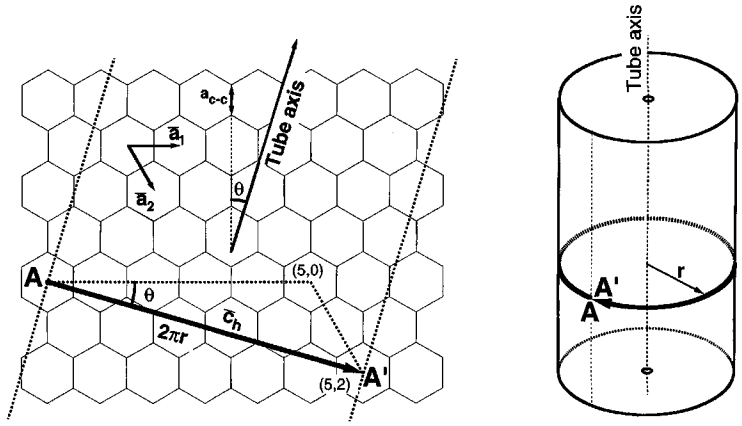


Figure 53. Schematic representation of the relation between r, θ, and (n, m) for a cylindrical carbon nanotube. Wrapping up of the graphene sheet in such a way that A coincides with A′ leads to the formation of a tube.

$\boldsymbol{c}_h = n\boldsymbol{a}_1 + m\boldsymbol{a}_2$ (Fig. 53) [178, 220]. The Hamada indices are generally used for theoretical work, while the radius r and the chiral angle θ can be determined experimentally. The radius r of a tube with Hamada indices (n, m) is given by the following relation

$$2r = \frac{\sqrt{3}}{\pi} a_{C-C} \sqrt{m^2 + mn + n^2} \tag{12}$$

and the chiral angle θ can be found from

$$\tan\theta = \frac{\sqrt{3}n}{2m + n} \tag{13}$$

with $a_{C-C} = 0.142\,\text{Å}$ (0.0142 nm), the nearest neighbor carbon–carbon distance [178]. Conversely, the measurement of r and θ for a monochiral tube allows m and n to be obtained from the relations

$$n = 2\sqrt{\frac{x}{y^2 + 3}} \tag{14}$$

and

$$m = \sqrt{\left(\frac{x}{y^2 + 3}\right)}(y - 1) \tag{15}$$

with

$$x = \frac{4r^2\pi^2}{3a_{C-C}^2} \tag{16}$$

and

$$y = \frac{\sqrt{3}}{\tan\theta} \tag{17}$$

11.7.4 Appendix D: Intuitive Direct Space Model for the Origin of Streaking in the Diffraction Patterns of Cylindrical Tubules

As far as diffraction is concerned, a tubule can be modeled as represented in Fig. 43, where the volume elements at a tangent to the tubes contain either a single graphene sheet (in monoshell tubes) or several random phase, disordered sheets. Within these sheets the tangential repeat distance along the normal to the cylinder axis is assumed to be a. It is thus possible to model the main diffraction effects due to the cylindrical curvature by considering each of these volume elements as a planar grating tangential to the cylinder and consisting of an array of equidistant, infinitely long scattering lines with separation a parallel to the tube axis. The problem is then reduced to two dimensions, as shown in Fig. 54a, representing a section

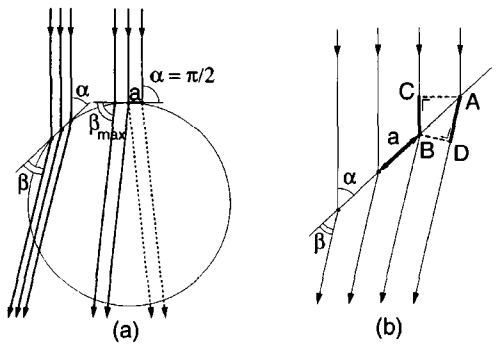


Figure 54. Direct space construction explaining the streaking of *hk*.0 spots in the diffraction patterns of carbon nanotubes. (a) Model for the circular monoshell tube. (b) Illustrating the derivation of the diffraction condition for a planar grating of parallel lines.

perpendicular to the tube axis. If we consider one such position of the grating (Fig. 54b), the diffraction condition expresses that intensity maxima occur if the path difference Δ of waves diffracted by successive 'lines' is an integer number of wavelengths λ

$$\Delta \equiv \mathrm{AD} - \mathrm{BC} = n\lambda$$

or

$$a(\cos\beta - \cos\alpha) = n\lambda \tag{18}$$

which is simply one of the three celebrated von Laue conditions. Under normal incidence ($\alpha = \pi/2$), the angle β is maximum, i.e., $\beta = \beta_m$, which is constant for a given reflection [$\beta_m = \arccos(n\lambda/a)$]. Hence Eq. (18) can be written as

$$\cos\beta - \cos\beta_m = \cos\alpha \tag{19}$$

Since $(\beta - \beta_m)$ is quite small, Eq. (19) can be rewritten as

$$\cos\alpha = (\beta_m - \beta)\sin\beta_m \tag{20}$$

or finally

$$\beta = \beta_m - \frac{\cos\alpha}{\sin\beta_m} \tag{21}$$

It is clear that as α decreases ($\cos\alpha$ increases), β also decreases for constant β_m, i.e., for a given reflection. At the same time, the intensity decreases with decreasing β since the incident intensity is proportional to $\sin\alpha$ and since, moreover, the atomic scattering factor also decreases with increasing scattering angle, i.e., with decreasing β. The reflection thus becomes elongated, fading out towards higher spatial frequencies. The interference between the beams diffracted by the two walls causes a quasiperiodic variation of the intensity, which is described exactly by Bessel functions [221, 222].

The smallest angle, α_{lim}, for which a diffraction maximum is still possible, is determined by the requirement that $\cos\beta \leqslant 1$, i.e., by $\cos\alpha_{lim} + \cos\beta = 1$ [from Eq. (19) or by $\cos\alpha_{lim} = 1 - \cos\beta_m$]. For this α value, β is zero and the diffracted beam is parallel to the planar diffraction grating. For smaller values of α some kind of 'total reflection' occurs, possibly contributing to the streak as well.

It is also worth noting that for small values of α (i.e., for grazing incidence), refraction of the electron beam occurs according to Snell's law. The latter phenomenon leads to observable features for *c*-reflections in the diffraction pattern of multishell tubules. The index of refraction being very close to one, the effect results in a small elongation of the $000l$ reflections along c^*.

An exact kinematical diffraction theory for the diffraction by helically arranged diffraction centers in single shell and multishell cylindrical tubules is formulated in the literature [218, 219].

11.7.5 Appendix E: Reciprocal Space of Helically Coiled Nanotubules[1]

Let us consider a coiled multilayer tubule with inner radius r of which the guiding circular helix is given in Cartesian coordinates by $x = R\cos\theta$, $y = R\sin\theta$, $z = (p/2\pi)\theta$ (R is the radius of the cylinder on which the helix is wound, p the pitch of the helix, and θ is the azimuth of the considered point); moreover, $\tan\phi_0 = (p/2\pi R)$, where ϕ_0 is the slope of the helix.

The geometry of the diffraction space of such a tubule (i.e., the loci of the reciprocal lattice nodes) can be obtained by considering the helical tube as consisting of short, straight, finite segments of tube of which the axis is always at a tangent to the guiding helix and is thus systematically varying in direction. The geometrical locus of the local axis of all such segments is a cone of revolution with semi-apex angle ϕ_0, with its apex in the origin and its axis parallel to the z-axis.

The geometry of diffraction space is now assumed to be the superposition of the loci contributed by all such elementary segments. The latter loci are approximated by those discussed above for infinitely long, straight tubules. In particular the loci of the $0\,0\,0\,l$ nodes, which are the prominent features, are circles centered on the origin of reciprocal space O_r with radii lc^* and situated in planes normal to the axis of the considered segment. The set of such circles in planes normal to the generators of the above-defined cone completely fill a spherical band limited by parallel circles on both sides of the equator of the sphere with radius lc^* centered on O_r; it extends in latitude from $-\phi_0$ to $+\phi_0$.

If the helix is not circular, but 'polygonized', the elementary segments are connected by bends caused by the presence of a pentagon–heptagon pair in diametrically opposite positions. In the diffraction space, the equatorial band now breaks up in discrete circles. The diffraction pattern obtained by intersecting diffraction space with Ewald's sphere (a plane) then consists of two $0\,0\,0\,l$ arcs broken up in discrete spots of which the density increases towards the extremities (Fig. 50d). The angular range of the arcs, as seen from the origin, is $2\phi_0$ at normal incidence of the electron beam, but it becomes larger with inclined incidence. For a full discussion we refer to Bernaerts et al. [159].

11.7.6 Appendix F: Semi-Apex Angles of Conical Tubes

Equation (9) determines a number of discrete semi-apex angles ϕ_0, which allow the formation of seamless conical tubes; this will be the case if $\theta = k\,60°$ (k: integer), i.e., for $\phi_0 = \arcsin(k/6) = 9.59°$, $19.47°$, $30°$, $41.81°$, $56.44°$, and $90°$ corresponding to $k = 1$–6. The acute extremity of such a seamless conical tube with a semi-apex angle ϕ_0 belonging to this series of special values [Eq. (9)] can be closed by means of a fullerene cap containing the adequate number N of pentagons. This number is given by $N = 6(1 - k/6)$. For instance, $k = 1$, i.e., the sharpest semi-apex angle

[1] Bernaerts et al. [159].

of 9.59° corresponds to a cap with $N = 5$ pentagons, whereas with $k = 5$, corresponding to the bluntest cone, a single pentagon ($N = 1$) is sufficient.

In Fig. 39b the measured semi-apex angle is, within experimental error, equal to the theoretical value of 9.59°. The cone is formed because of the presence in A of one pentagon (in each sheet!); the configuration is therefore asymmetrical. For closure of the cone, caps containing five pentagons are required in each of the concentric sheets.

Seamless cones (and also low energy scrolls) can only be formed for the specific angles given above. In Fig. 52 the semi-apex angle is ~70°, which is not one of these specific angles; as a result, misfit occurs between successive sheets. Nevertheless, the nearest ideal case corresponds to $k = 5$, i.e., one pentagon is required along the axis [40].

The pentagons in the caps are whenever possible arranged so as to be all 'isolated', i.e., surrounded by hexagons. This pentagon isolation rule is fulfilled in the cage fullerenes such as C_{60} and C_{70}.

Acknowledgements

The authors thank numerous colleagues for the use of photographs, either unpublished ones or from joint papers with the present authors. Reference to the original publications is included in the figure captions.

We also acknowledge illuminating discussions with Prof. A. Lucas (Namur, Belgium). D. Bernaerts thanks the Belgian National Fund for Scientific Research for financial support. This text presents research results of the Belgian Programme on Inter-University Poles of Attraction initiated by the Belgian State Prime Minister's Office of Science Policy Programming. Scientific responsibility is assumed by the authors.

11.8 References

[1] H. Marsh, *Introduction to Carbon Science*, Butterworths, London **1989**.

[2] K. Lonsdale, *Proc. R. Soc. London* **1941**, *A179*, 315.

[3] J. Wilks, E. Wilks, *Properties and Applications of Diamond*, Butterworth-Heinemann, Oxford **1991**.

[4] B. V. Spitsyn, L. L. Bouilov, B. V. Derjaguin, *J. Cryst. Growth* **1981**, *52*, 219.

[5] S. Matsumoto, Y. Saito, M. Tsutsumi, N. Setaka, *J. Mater. Sci.* **1982**, *17*, 3106.

[6] S. Matsumoto, M. Hino, T. Kobayashi, *Appl. Phys. Lett.* **1987**, *51*, 737.

[7] K. Kurihara, K. Sasaki, M. Kawarada, N. Koshino, *Appl. Phys. Lett.* **1988**, *52*, 437.

[8] H. W. Kroto, J. R. Heith, S. C. O'Brien, R. F. Curl, R. E. Smalley, *Nature* **1985**, *318*, 162.

[9] W. Krätschmer, L. D. Lamb, K. Fostiropoulos, D. R. Huffman, *Nature* **1990**, *347*, 354.

[10] S. M. Gorum, K. M. Creegan, R. D. Sherwood, D. M. Cox, V. W. Day, C. S. Day, R. M. Upton, C. E. Briant, *J. Chem. Soc., Chem. Commun.* **1991**, 1556.

[11] G. Van Tendeloo, M. Op de Beeck, S. Amelinckx, J. Bohr, W. Krätschmer, *Europhys. Lett.* **1991**, *15*, 295.

[12] J. Bohr, D. Gibbs, S. K. Sinha, W. Krätschmer, G. Van Tendeloo, E. Larsen, H. Egsgaard, L. E. Berman, *Europhys. Lett.* **1992**, *17*, 327.

[13] Z. G. Li, P. J. Fagan, *Chem. Phys. Lett.* **1992**, *194*, 461.

[14] G. Van Tendeloo, C. Van Heurck, J. Van Landuyt, S. Amelinckx, M. A. Verheijen, P. H. M. van Loosdrecht, G. Meijer, *J. Phys. Chem.* **1992**, *96*, 7424.

[15] H.-B. Bürgi, E. Blanc, D. Schwarzenbach, S. Liu, Y.-J. Lu, M. M. Kappes, J. A. Ibers, *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 640.

[16] W. I. F. David, R. M. Ibberson, J. C. Matthewman, K. Prassides, T. J. S. Dennis, J. P. Hare, H. W. Kroto, R. Taylor, D. R. M. Walton, *Nature* **1991**, *353*, 147.

[17] P. A. Heiney, J. E. Fisher, A. R. McGhie, W. J. Romanow, A. M. Denestein, J. P. McCauley, A. B. Smith III, *Phys. Rev. Lett.* **1991**, *66*, 2911.

[18] Y. Z. Li, M. Chander, J. C. Patrin, J. H. Weaver, L. P. F. Chibante, R. Smalley, *Science* **1991**, *253*, 429.

[19] Y. Z. Li, J. C. Patrin, M. Chander, J. H. Weaver, L. P. F. Chibante, R. E. Smalley, *Science* **1991**, *252*, 547.

[20] H. P. Lang, V. Thommen-Geiser, C. Bohm, M. Felder, J. Frommer, R. Wiesendanger, H. Werner, R. Schlögl, A. Zahab, P. Bernier, G. Gerth, D. Anselmetti, H.-J. Güntherodt, *Appl. Phys.* **1993,** *A56*, 197.
[21] T. Hashizume, K. Motai, X. D. Wang, H. Shinohara, T. Sakurai, *Sci. Rep. RITU* **1993,** *A39*, 51.
[22] Y. Z. Li, J. C. Patrin, M. Chander, J. H. Weaver, K. Kikuchi, Y. Achiba, *Phys. Rev.* **1993,** *B47*, 10867.
[23] M. A. Verheijen, M. S. Couto, K. W. M. Koutstaal, W. J. P. van Enckevort, *Phil. Mag.* **1995,** *A72*, 1141.
[24] C. N. R. Rao, A. Gouvindaraj, H. N. Aiyer, R. Seshadri, *J. Phys. Chem.* **1995,** *99*, 16814.
[25] B. Keita, L. Nadjo, R. Céolin, V. Agafonov, D. André, H. Szwarc, J. Dugué, C. Fabre, A. Rassat, *Chem. Phys.* **1994,** *179*, 595.
[26] J. E. Fisher, P. A. Heiney, *J. Phys. Chem. Solids* **1993,** *54*, 1725.
[27] S. Iijima, *Nature* **1991,** *354*, 56.
[28] T. W. Ebbesen, *Annu. Rev. Mater. Sci.* **1994,** *24*, 235.
[29] R. E. Smalley in *Proc. of the Robert A. Welch Foundation Conference on Chemical Research XXXVI*, Houston, TX, 26–27 October **1992,** p. 161.
[30] P. M. Ajayan, *Condensed Matter News* **1995,** *4*, 9.
[31] O. Zhou, R. M. Flemming, D. W. Murphy, C. H. Chen, R. C. Haddon, A. P. Ramirez, S. H. Glarum, *Science* **1994,** *263*, 1744.
[32] S. Amelinckx, D. Bernaerts, X. B. Zhang, G. Van Tendeloo, J. Van Landuyt, *Science* **1995,** *267*, 1334.
[33] S. Iijima, T. Ichihashi, *Nature* **1993,** *363*, 603.
[34] D. S. Bethune, C. H. Kiang, M. S. de Vries, G. Gorman, R. Savoy, J. Vazquez, R. Beyers, *Nature* **1993,** *363*, 605.
[35] P. M. Ajayan, J. M. Lambert, P. Bernier, L. Barbedette, C. Colliex, J. M. Planeix, *Chem. Phys. Lett.* **1993,** *215*, 509.
[36] T. Guo, P. Nikolaev, A. Thess, D. T. Colbert, R. E. Smalley, *Chem. Phys. Lett.* **1995,** *243*, 49.
[37] A. Thess, R. Lee, P. Nikolaev, H. Dai, P. Petit, J. Robert, C. Xu, Y. H. Lee, S. G. Kim, D. T. Colbert, G. Scuseria, D. Tomanek, J. E. Fisher, R. E. Smalley, submitted to *Science*.
[38] X. B. Zhang, X. F. Zhang, D. Bernaerts, G. Van Tendeloo, S. Amelinckx, J. Van Landuyt, V. Ivanov, J. B. Nagy, P. Lambin, A. A. Lucas, *Europhys. Lett.* **1994,** *27*, 141.
[39] V. Ivanov, J. B. Nagy, P. Lambin, A. Lucas, X. B. Zhang, X. F. Zhang, D. Bernaerts, G. Van Tendeloo, J. Van Landuyt, *Chem. Phys. Lett.* **1994,** *223*, 329.
[40] S. Amelinckx, W. Luyten, T. Krekels, G. Van Tendeloo, J. Van Landuyt, *J. Cryst. Growth* **1992,** *121*, 543.
[41] W. Luyten, T. Krekels, S. Amelinckx, G. Van Tendeloo, D. Van Dyck, J. Van Landuyt, *Ultramicroscopy* **1993,** *49*, 123.
[42] S. Amelinckx, P. Delavignette, M. Heerschap in *Chemistry and Physics of Carbon, Vol. 1* (Ed.: P. L. Walker, Jr.), Marcel Dekker, New York **1965,** pp. 1–71.
[43] P. Delavignette, S. Amelinckx, *J. Nucl. Mater.* **1962,** *5*, 17.
[44] S. Amelinckx, P. Delavignette, *J. Appl. Phys.* **1960,** *31*, 2126; P. Delavignette, S. Amelinckx, *J. Appl. Phys.* **1960,** *31*, 1691.
[45] M. J. Whelan, *Proc. R. Soc. London, Ser. A* **1958,** *249*, 114.
[46] R. Siems, S. Amelinckx, P. Delavignette, *Z. Physik* **1961,** *165*, 502.
[47] M. Heerschap, P. Delavignette, *Appl. Phys. Lett.* **1963,** *2*, 229.
[48] M. Heerschap, P. Delavignette, S. Amelinckx, *Carbon* **1964,** *1*, 235.
[49] S. Tolansky in *Multiple Beam Interferometry*, Oxford University Press, Oxford **1948.**
[50] J. M. Thomas in *Chemistry and Physics of Carbon, Vol. I* (Ed.: P. L. Walker, Jr.), Marcel Dekker, New York **1965,** pp. 121–202.
[51] G. R. Hennig, *Z. Elektrochem.* **1962,** *66*, 629.
[52] G. R. Hennig, *J. Inorg. Nucl. Chem.* **1962,** *24*, 1129.
[53] F. A. Horn, *Nature* **1952,** *170*, 581.
[54] J. M. Thomas, C. Roscoe in *Chemistry and Physics of Carbon, Vol. 3* (Ed.: P. L. Walker, Jr.), Marcel Dekker, New York **1968,** pp. 1–44.
[55] G. Binnig, H. Fuchs, C. Gerber, E. Stoll, E. Tosatti, *Europhys. Lett.* **1986,** *1*, 31.
[56] D. Tomanek, S. G. Louie, *Phys. Rev.* **1988,** *B37*, 8327.
[57] H. J. Mamin, E. Ganz, D. W. Abraham, R. E. Thomson, J. Clarke, *Phys. Rev.* **1996,** *B34*, 9010.
[58] S. P. Kelty, C. M. Lieber, *J. Phys. Chem.* **1989,** *93*, 5983.
[59] M. Omar, N. S. Pandya, S. Tolansky, *Proc. R. Soc.* **1954,** *A225*, 33.
[60] F. C. Franck, K. E. Puttick, *Phil. Mag.* **1958,** *3*, 1273.
[61] S. Tolansky in *Physical Properties of Diamond* (Ed.: R. Berman), Clarendon, Oxford **1963,** pp. 135–173.
[62] E. M. Wilks, J. Wilks in *Diamond Research 1978*, supplement to *Industrial Diamond Review* **1978,** pp. 2–10.
[63] A. R. Lang in *Modern Diffraction and Imaging Techniques in Materials Science* (Eds.: S. Amelinckx, R. Gevers, G. Remant, J. Van Landuyt), North Holland, Amsterdam **1970,** pp. 407–479.

[64] B. K. Tanner, *X-Ray Diffraction Topography*, Pergamon, Oxford **1976**.
[65] P. Pirouz, D. J. H. Cockayne, N. Sumida, P. Hirsh, A. R. Lang, *Proc. R. Soc. London* **1983,** *A386*, 241.
[66] K. Suzuki, M. Ichihara, S. Takeuchi, N. Ohtake, M. Yoshikawa, K. Hirabayashi, N. Kurihara, *Phil. Mag.* **1992,** *A65*, 657.
[67] T. Evans, C. Phaal, *Proc. R. Soc. London* **1962,** *A270*, 538.
[68] G. S. Woods, *Phil. Mag.* **1976,** *34*, 993.
[69] A. R. Lang in *The Properties of Diamond* (Ed.: J. E. Field), Academic, London **1979,** pp. 425–469.
[70] L. A. Bursil, J. L. Hutchinson, J. L. Lang, N. Sumida, *Nature* **1981,** *292*, 518.
[71] J. C. Barry, L. A. Bursil, J. L. Hutchinson, *Phil. Mag.* **1983,** *A48*, 109.
[72] P. Humble, *Proc. R. Soc. London* **1982,** *A381*, 65.
[73] J. C. Barry, *Phil. Mag.* **1991,** *A64*, 111.
[74] R. F. Stephenson, Ph.D. Thesis, University of Reading, Reading **1978**.
[75] R. Burt, Ph.D. Thesis, University of Reading, Reading, **1980**.
[76] J. Maguire, Ph.D. Thesis, University of Reading, Reading **1983**.
[77] P. B. Hirsch, J. L. Hutchison, J. Titchmarsh, *Phil. Mag. Lett.* **1986,** *A54*, L49.
[78] J. C. Barry, *Ultramicroscopy* **1986,** *20*, 169.
[79] G. S. Woods, *Proc. R. Soc.* **1986,** *A407*, 219.
[80] G. Van Tendeloo, W. Luyten, G. S. Woods, *Phil. Mag. Lett.* **1990,** *61*, 343.
[81] W. Luyten, G. Van Tendeloo, P. J. Fallon, G. S. Woods, *Phil. Mag.* **1994,** *A69*, 767.
[82] J. I. Goldstein, H. Yakowitz, *Practical Scanning Electron Microscopy*, Plenum, New York **1975**.
[83] B. E. Williams, H. S. Kong, J. T. Glass, *J. Mater. Res.* **1990,** *5*, 801.
[84] J. Narayan, *J. Mater. Res.* **1990,** *5*, 2414.
[85] J. C. Angus, M. Sunkara, S. R. Sahaida, J. T. Glass, *J. Mater. Res.* **1992,** *7*, 3001.
[86] W. Marciniak, K. Fabisiak, S. Orzeszko, F. Rozploch, *J. Cryst. Growth* **1992,** *123*, 587.
[87] D. Shechtman, J. L. Hutchison, L. H. Robins, E. N. Farabaugh, A. Feldman, *J. Mater. Res.* **1993,** *8*, 473.
[88] D. Shechtman, A. Feldman, M. D. Vaudin, J. L. Hutchison, *Appl. Phys. Lett.* **1993,** *62*, 487.
[89] N. Jiang, Z. Zhang, B. W. Sun, D. Shi, *Appl. Phys. Lett.* **1993,** *63*, 328.
[90] Q. J. Gao, H. Yuan, X. Zhang, X. B. Fen, J. H. Li, Y. C. Yang, Z. D. Lin, X. F. Pen, B. X. Yang, G. S. Jiang, F. X. Lu, *Thin Solid Films* **1992,** *221*, 34.
[91] H. Kawarada, T. Suesada, H. Nagasawa, *Appl. Phys. Lett.* **1995,** *66*, 583.
[92] K.-A. Feng, J. Yiang, Z. Lin, *Phys. Rev.* **1995,** *B 51*, 2264.
[93] F. R. Sivazlian, J. T. Glass, B. R. Stoner, *J. Mater. Res.* **1994,** *9*, 2487.
[94] N. Jiang, B. W. Sun, Z. Zhang, Z. Lin, *J. Mater. Res.* **1994,** *9*, 2695.
[95] C. L. Jia, K. Urban, X. Jiang, *Phys. Rev.* **1995,** *B 52*, 5164.
[96] Y. Tzou, J. Bruley, F. Ernst, M. Rühle, R. Raj, *J. Mater. Res.* **1994,** *9*, 1566.
[97] W. Luyten, G. Van Tendeloo, S. Amelinckx, J. L. Collins, *Phil. Mag.* **1992,** *A 66*, 899.
[98] H. Bender, A. De Veirman, J. Van Landuyt, S. Amelinckx, *Appl. Phys.* **1986,** *A 39*, 83.
[99] J. E. Fisher, P. A. Heiney, A. R. McGhie, W. J. Romanow, A. M. Denenstein, J. P. McCauley Jr., A. B. Smith III, *Science* **1991,** *252*, 1288.
[100] S. Amelinckx, C. Van Heurck, D. Van Dyck, G. Van Tendeloo, *Phys. Status Solidi (a)* **1992,** *131*, 589.
[101] S. J. Duclos, K. Brister, R. C. Haddon, A. R. Kortan, F. A. Thiel, *Nature* **1991,** *351*, 380.
[102] W. I. F. David, R. M. Ibberson, T. J. S. Dennis, J. P. Hare, K. Prassides, *Europhys. Lett.* **1992,** *18*, 219.
[103] G. Van Tendeloo, S. Amelinckx, M. A. Verheijen, P. H. M. Loosdrecht, G. Meijer, *Ultramicroscopy* **1993,** *51*, 168.
[104] G. Van Tendeloo, S. Amelinckx, M. A. Verheijen, P. H. M. van Loosdrecht, G. Meijer, *Phys. Rev. Lett.* **1992,** *69*, 1065.
[105] J. E. Fisher, D. E. Luzzi, K. Kniaz, A. R. McGhie, D. A. Ricketts-Foot, W. R. Romanow, G. B. M. Vaughan, P. A. Heiney, D. Li, A. L. Smith, R. M. Strongin, M. A. Cichy, L. Brard, A. B. Smith III, *Phys. Rev.* **1993,** *B 47*, 14614.
[106] F. Banhart, M. Forster, W. Krätschmer, H.-E. Schaefer, *Phil. Mag. Lett.* **1992,** *65*, 283.
[107] S. Muto, G. Van Tendeloo, S. Amelinckx, *Phil. Mag.* **1993,** *B 67*, 443.
[108] M. A. Verheijen, W. J. P. van Enckevort, G. Meijer, *Chem. Phys. Lett.* **1993,** *216*, 72.
[109] M. A. Verheijen, H. Meekes, G. Meijer, P. Bennema, J. L. de Boer, S. van Smaalen, G. Van Tendeloo, S. Amelinckx, S. Muto, J. van Landuyt, *Chem. Phys.* **1992,** *166*, 287.
[110] G. Van Tendeloo, S. Amelinckx, J. L. De Boer, S. Van Smaalen, M. A. Verheijen, H. Meekes, G. Meijer, *Europhys. Lett.* **1993,** *21*, 329.
[111] G. B. M. Vaughan, P. A. Heiney, J. E. Fisher, D. E. Luzzi, D. A. Ricketts-Foot, A. R. McGhie, Y.-W. Hui, A. L. Smith, D. E. Cox, W. J. Romanow, B. H. Allen, N. Coustel, J. P. McCauley, Jr., A. B. Smith III, *Science* **1991,** *254*, 1350.

[112] C. Christides, I. M. Thomas, T. J. S. Dennis, K. Prassides, *Europhys. Lett.* **1993,** *22*, 611.
[113] C. Meingast, F. Gugenberger, M. Haluska, H. Kuzmany, G. Roth, *Appl. Phys.* **1993,** *A 56*, 227.
[114] M. C. Valsakumar, N. Subramanian, M. Yousuf, P. C. Sahu, Y. Hariharan, A. Bharathi, V. Sankara Sastry, J. Janaki, G. V. N. Rao, T. S. Radhakrishnan, C. S. Sundar, *Phys. Rev.* **1993,** *B 48*, 9080.
[115] G. B. M. Vaughan, P. A. Heiney, D. E. Cox, J. E. Fisher, A. R. McGhie, A. L. Smith, R. M. Strongin, M. A. Cichy, A. B. Smith III, *Chem. Phys.* **1993,** *178*, 599.
[116] E. Blanc, H. B. Burgi, R. Restori, D. Schwarzenback, P. Stellberg, P. Venugopalan, *Europhys. Lett.* **1994,** *27*, 359.
[117] S. K. Ramasesha, A. K. Singh, R. Seshadri, A. K. Sood, C. N. R. Rao, *Chem. Phys. Lett.* **1994,** *220*, 203.
[118] C. Christides, T. J. Dennis, K. Prasides, R. L. Cappelletti, D. A. Neumann, J. R. D. Copley, *Phys. Rev.* **1994,** *B 49*, 2897.
[119] R. Blinc, J. Dolinsek, J. Seliger, D. Arcon, *Solid State Commun.* **1993,** *88*, 9.
[120] R. Tycko, G. Dabbagh, G. B. M. Vaughan, P. A. Heiney, R. M. Strongin, M. A. Cichy, A. B. Smith, *J. Chem. Phys.* **1993,** *99*, 7554.
[121] A. R. McGhie, J. E. Fisher, P. A. Heiney, P. W. Stephens, R. L. Cappelletti, D. A. Neumann, W. H. Mueller, H. Mohn, H.-U. ter Meer, *Phys. Rev.* **1994,** *B 49*, 12614.
[122] X. B. Zhang, X. F. Zhang, S. Amelinckx, H. Werner, *Appl. Phys.* **1994,** *A 58*, 107.
[123] X. F. Zhang, X. B. Zhang, D. Bernaerts, G. Van Tendeloo, S. Amelinckx, J. Van Landuyt, H. Werner, *Ultramicroscopy* **1994,** *55*, 25.
[124] Y. Z. Li, J. C. Patrin, M. Chander, J. H. Weaver, L. P. F. Chibante, R. E. Smalley, *Phys. Rev.* **1992,** *B 46*, 12914.
[125] S. Behler, H. P. Lang, S. H. Pan, V. Thommen-Geiser, R. Hofer, M. Bernasconi, H.-J. Güntherodt in *Springer Series in Solid-State Sciences, Vol. 117, Electronic Properties of Fullerenes* (Eds.: H. Kuzmany, J. Fink, M. Mehring, S. Roth), Springer, Berlin **1993.**
[126] R. Lüthi, H. Haefke, E. Meyer, L. Howald, H.-P. Lang, G. Gerth, H.-J. Güntherodt, *Z. Phys.* **1994,** *B 95*, 1.
[127] H. N. Aiyer, A. Govindaraj, C. N. R. Rao, *Phil. Mag. Lett.* **1995,** *72*, 185.
[128] Y. Saito, Y. Ishikawa, A. Ohshita, H. Shinohara, H. Nagashima, *Phys. Rev.* **1992,** *B 46*, 1846.
[129] Y. Saito, T. Yoshikawa, N. Fujimoto, H. Shinohara, *Phys. Rev.* **1993,** *B 48*, 9182.
[130] W. L. Zhou, W. Zhao, K. K. Fung, L. Q. Chen, Z. B. Zhang, *Physica* **1993,** *C 214*, 19.
[131] I. Rusakova, A. Hamed, P. H. Hor, *J. Mater. Res.* **1994,** *9*, 2814.
[132] Y. Takahashi, *Jpn. J. Appl. Phys.* **1994,** *33*, 4104.
[133] Y. Takahashi, K. Hayashi, *J. Electron Microsc.* **1994,** *43*, 378.
[134] W. B. Zhao, X. D. Zhang, Z. Y. Ye, J. L. Zhang, C. Y. Li, D. L. Yin, Z. N. Gu, X. H. Zhou, Z. X. Jin, *Thin Solid Films* **1994,** *240*, 14.
[135] Y. Saito, T. Yoshikawa, Y. Ishikawa, H. Nagashima, H. Shinohara, *Mater. Sci. Eng.* **1993,** *B 19*, 18.
[136] D. Bernaerts, G. Van Tendeloo, S. Amelinckx, K. Hevesi, G. Gensterblum, L. M. Yu, J.-J. Pireaux, F. Grey, J. Bohr, accepted for publication in *J. Appl. Phys.*
[137] G. Gensterblum, L.-M. Yu, J.-J. Pireaux, P. A. Thiry, R. Caudano, J.-M. Themlin, S. Bouzidi, F. Coletti, J.-M. Debever, *Appl. Phys.* **1993,** *A 56*, 175.
[138] G. Gensterblum, K. Hevesi, B.-Y. Han, L.-M. Yu, D. Bernaerts, G. Van Tendeloo, S. Amelinckx, G. Bendele, T. Buslaps, M. Foss, R. Feidenhans'l, R. L. Johnson, J.-J. Pireaux, P. A. Thiry, R. Caudano, G. Le Lay, *Phys. Rev.* **1994,** *B 50*, 11981.
[139] S. Iijima, *J. Cryst. Growth* **1980,** *50*, 675.
[140] Y. Saito, T. Yoshikawa, M. Inagaki, M. Tomita, T. Hayashi, *Chem. Phys. Lett.* **1993,** *204*, 277.
[141] Y. Ando, S. Iijima, *Jpn. J. Appl. Phys.* **1993,** *32*, 107.
[142] G. H. Taylor, J. D. Fitzgerald, L. Pang, M. A. Wilson, *J. Cryst. Growth* **1994,** *135*, 157.
[143] D. T. Colbert, J. Zhang, S. M. McClure, P. Nikolaev, Z. Chen, J. H. Hafner, D. W. Owens, P. G. Kotula, C. B. Carter, J. H. Weaver, A. G. Rinzler, R. E. Smalley, *Science* **1994,** *266*, 1218.
[144] H. Hiura, T. W. Ebbesen, J. Fujita, K. Tanigaki, T. Takada, *Nature* **1994,** *367*, 148.
[145] S. C. Tsang, P. de Oliveira, J. J. Davis, M. L. H. Green, H. A. O. Hill, *Chem. Phys. Lett.* **1996,** *249*, 413.
[146] M. Ge, K. Sattler, *Science* **1993,** *260*, 515.
[147] M. Ge, K. Sattler, *Appl. Phys. Lett.* **1994,** *65*, 2284.
[148] S. Iijima, P. M. Ajayan, T. Ichihashi, *Phys. Rev. Lett.* **1992,** *69*, 3100.
[149] S. Iijima, T. Ichihashi, Y. Ando, *Nature* **1992,** *356*, 776.
[150] H. Jaeger, V. W. Malsen, A. McL. Mathieson, *Carbon* **1992,** *30*, 269.
[151] P. M. Ajayan, S. Iijima, *Nature* **1992,** *358*, 23.
[152] P. M. Ajayan, T. Ichihashi, S. Iijima, *Chem. Phys. Lett.* **1993,** *202*, 384.

[153] X. F. Zhang, X. B. Zhang, G. Van Tendeloo, S. Amelinckx, M. Op de Beeck, J. Van Landuyt, *J. Cryst. Growth* **1993,** *130*, 368.
[154] V. P. Dravid, X. Lin, Y. Wang, X. K. Wang, A. Yee, J. B. Ketterson, R. P. H. Chang, *Science* **1993,** *259*, 1601.
[155] X. B. Zhang, X. F. Zhang, S. Amelinckx, G. Van Tendeloo, J. Van Landuyt, *Ultramicroscopy* **1994,** *54*, 237.
[156] M. Liu, J. M. Cowley, *Carbon* **1994,** *32*, 393.
[157] M. Liu, J. M. Cowley, *Mater. Sci. Eng.* **1994,** *A 185*, 131.
[158] M. Liu, J. M. Cowley, *Ultramicroscopy* **1994,** *53*, 333.
[159] D. Bernaerts, X. B. Zhang, X. F. Zhang, S. Amelinckx, G. Van Tendeloo, J. Van Landuyt, V. Ivanov, J. B. Nagy, *Phil. Mag.* **1995,** *A 71*, 605.
[160] T. W. Ebbesen, P. M. Ajayan, *Nature* **1992,** *358*, 220.
[161] T. W. Ebbesen, P. M. Ajayan, H. Hiura, K. Tanigaki, *Nature* **1994,** *367*, 519.
[162] M. J. Yacaman, M. Miki-Yoshida, L. Rendon, J. G. Santiesteban, *Appl. Phys. Lett.* **1993,** *62*, 657.
[163] V. Ivanov, A. Fonseca, J. B. Nagy, A. Lucas, P. Lambin, D. Bernaerts, X. B. Zhang, *Carbon* **1995,** *33*, 1727.
[164] R. Bacon, *J. Appl. Phys.* **1960,** *31*, 283.
[165] A. Oberlin, M. Endo, T. Koyana, *J. Cryst. Growth* **1976,** *32*, 335.
[166] R. T. K. Baker, *Carbon* **1989,** *27*, 315.
[167] S. Motojima, M. Kawaguchi, K. Nozaki, H. Iwanaga, *Carbon* **1991,** *29*, 379.
[168] N. M. Rodriguez, *J. Mater. Res.* **1993,** *8*, 3233.
[169] P. M. Ajayan, S. Iijima, *Nature* **1993,** *361*, 333.
[170] P. M. Ajayan, T. W. Ebbesen, T. Ichihashi, S. Iijima, K. Tanigaki, H. Hiura, *Nature* **1993,** *362*, 522.
[171] S. Seraphin, D. Zhou, J. Jiao, J. C. Withers, R. Loutfy, *Nature* **1993,** *362*, 503.
[172] P. M. Ajayan, C. Colliex, J. M. Lambert, P. Bernier, L. Barbedette, M. Tencé, O. Stephan, *Phys. Rev. Lett.* **1994,** *72*, 1722.
[173] X. F. Zhang, X. B. Zhang, G. Van Tendeloo, G. Meijer, *Carbon* **1994,** *32*, 363.
[174] S. C. Tsang, Y. K. Chen, P. J. F. Harris, M. L. H. Green, *Nature* **1994,** *372*, 159.
[175] H. Funasaka, K. Sugiyama, K. Yamamoto, T. Takahashi, *Chem. Phys. Lett.* **1995,** *236*, 277.
[176] P. M. Ajayan, O. Stephan, P. Redlich, C. Colliex, *Nature* **1995,** *375*, 564.
[177] M. S. Dresselhaus, G. Dresselhaus, R. Saito, *Phys. Rev.* **1992,** *B 45*, 6234.
[178] M. S. Dresselhaus, G. Dresselhaus, R. Saito, *Solid State Commun.* **1992,** *84*, 201.
[179] Y. Saito, T. Yoshikawa, S. Bandow, M. Tomita, T. Hayashi, *Phys. Rev.* **1993,** *B 48*, 1907.
[180] M. S. Zwanger, F. Banhart, *Phil. Mag.* **1995,** *B 72*, 149.
[181] X.-D. Fan, L. A. Bursil, *Phil. Mag.* **1995,** *A 72*, 139.
[182] S. Wang, D. Zhou, *Chem. Phys. Lett.* **1994,** *225*, 165.
[183] M. Bretz, G. Demczyk, L. Zhang, *J. Cryst. Growth* **1994,** *141*, 304.
[184] S. L. Cullen, C. B. Boothroyd, C. J. Humphreys, *Ultramicroscopy* **1994,** *56*, 127.
[185] K. Yada, *Acta Cryst.* **1967,** *23*, 704.
[186] D. Bernaerts, S. Amelinckx, G. Van Tendeloo, J. Van Landuyt, accepted for publication in *Journal of Crystal Growth.*
[187] S. Iijima in *ICEM* 13 July 17–22, Paris, **1994.**
[188] X. B. Zhang, S. Amelinckx, *Carbon* **1994,** *32*, 1537.
[189] B. Devouard, A. Baronnet, to be published in *Eur. J. Mineralogy.*
[189a] D. Bernaerts, M. Op de Beeck, S. Amelinkcx, J. Van Landuyt, G. Van Tendeloo, accepted for publication in *Phil. Mag. A.*
[190] S. Amelinckx, D. Bernaerts, G. Van Tendeloo, J. Van Landuyt, A. A. Lucas, M. Mathot, P. Lambin in *Physics and Chemistry of Fullerenes and Derivatives, Proc. Int. Winterschool on the Electronic Properties of Novel Materials* (Eds.: H. Kuzmany, J. Fink, M. Mehring, S. Roth), World Scientific, Singapore **1995,** p. 515.
[191] J. T. Finch, *J. Mol. Biol.* **1972,** *66*, 291.
[192] L. Margulis, P. Dluzewski, Y. Feldman, R. Tene, *J. Microsc.* **1996,** *181*, 68.
[193] S. Amelinckx in *The Direct Observation of Dislocations, Solid State Physics suppl. 6* (Eds.: F. Seitz, D. Turnbull), Academic, New York **1964.**
[194] D. Ugarte, *Nature* **1992,** *359*, 707.
[195] D. Ugarte, *Europhys. Lett.* **1993,** *22*, 45.
[196] D. Ugarte, *Chem. Phys. Lett.* **1994,** *207*, 473.
[197] D. Ugarte, W. A. de Heer in *Springer Series in Solid State Sciences 117: Electronic Properties of Fullerenes* (Eds.: H. Kuzmany, J. Fink, M. Mehring, S. Roth), Springer, Berlin, **1993.**
[198] M. Yoshida, E. Osawa, *Fullerene Sci. Technol.* **1993,** *1*, 55.
[199a] F. Banhart and P. M. Ajayan, *Nature* **1996,** *382*, 433.
[199b] L. C. Qin and S. Iijima, accepted for publication in *Chemical Physics Letters.*
[200] F. Banhart, P. M. Ajayan, P. Redlich, T. Füller, submitted for publication.
[201] D. N. Weldon, W. J. Blau, H. W. Zandbergen, *Chem. Phys. Lett.* **1995,** *241*, 365.

[202] S. Iijima, C. Brabec, A. Maiti, J. Bernholc, *J. Chem. Phys.* **1996,** *104*, 2089.
[203] R. S. Ruoff, J. Tersoff, D. C. Lorents, S. Subramoney, B. Chan, *Nature* **1993,** *364*, 514.
[204] N. G. Shopra, L. X. Benedict, V. H. Crespi, M. L. Cohen, S. G. Louie and A. Zettl, *Nature* **1995,** *377*, 135.
[205] M. M. J. Treacy, T. W. Ebbesen and J. M. Gibson, *Nature* **1996,** *381*, 676.
[206] S. Iijima, *Mater. Sci. Eng.* **1993,** *B 19*, 172.
[207] D. Zhou, S. Seraphin, *Chem. Phys. Lett.* **1995,** *238*, 286.
[208] B. I. Dunlap, *Phys. Rev.* **1992,** *B 46*, 1933.
[209] B. I. Dunlap, *Phys. Rev.* **1994,** *B 49*, 5643.
[210] P. Lambin, A. Fonseca, J. P. Vigneron, J. B. Nagy, A. A. Lucas, *Chem. Phys. Lett.* **1995,** *245*, 85.
[211] S. Ihara, S. Itoh, J. I. Kitakami, *Phys. Rev.* **1993,** *B 48*, 5643.
[212] S. Amelinckx, X. B. Zhang, D. Bernaerts, X. F. Zhang, V. Ivanov, J. B. Nagy, *Science* **1994,** *265*, 635.
[213] Z. G. Li, P. J. Fagan, L. Liang, *Chem. Phys. Lett.* **1993,** *207*, 148.
[214] L. A. Bursill, J.-L. Peng, X.-D. Fan, *Phil. Mag.* **1995,** *A 71*, 1161.
[215] G. Hu, X. F. Zhang, D. P. Yu, S. Q. Feng, W. Xu, Z. Zhang, *Solid State Commun.* **1996,** *98*, 547.
[216] P. J. F. Harris, M. L. H. Green, S. C. Tsang, *J. Chem. Soc. Faraday Trans.* **1993,** *89*, 1189.
[217] M. Ge, K. Sattler, *Chem. Phys. Lett.* **1994,** *220*, 192.
[218] Y. Saito, T. Yoshikawa, *J. Cryst. Growth* **1993,** *134*, 154.
[219] S. Amelinckx, B. Devouard, A. Baronet, accepted for publication in *Acta Cryst. A.*
[220] N. Hamada, S. Sawada, A. Oshiyama, *Phys. Rev. Lett.* **1992,** *68*, 1579.
[221] A. A. Lucas, V. Bruyninckx, P. Lambin, D. Bernaerts, S. Amelinckx, J. Van Landuyt, G. Van Tendeloo, in preparation.
[222] L. C. Qin, *J. Mater. Res.* **1994,** *9*, 2450.

Handbook of Microscopy
S. Amelinckx, D. van Dyck,J. van Landuyt , G. van Tendeloo

12 Composite Structural Materials

12.1 Introduction

By the term 'composites' we mean materials fabricated from two distinct constituent phases, which after fabrication have retained their chemical, physical, or geometrical identity. Examples are silicon carbide fibers in a silicon carbide matrix, silicon carbide particles in an aluminum alloy matrix, and carbon fibers in a polymer matrix. As illustrated by these examples, constituent phases may be a ceramic, a metal, or a polymer phase. The motivation for the development of these materials is to create materials with a unique combination of properties that cannot be found among the monolithic materials. These properties include physical (e.g. electronic, magnetic, and optical), chemical, or mechanical properties. Here we consider only composites intended for structural applications, where mechanical properties such as strength, stiffness, toughness, and creep resistance play an important role. Usually a composite consists of a contiguous phase, the 'matrix', and a discrete phase, the 'reinforcement phase', which may be composed of particles, platelets, short or long fibers. Composites are often classified according to the nature of the matrix phase and thus are described as metal matrix (MMC), polymer matrix (PMC), and ceramic matrix composites (CMC).

To understand composites fully, a knowledge of their microstructure is indispensable. The relative amount of the reinforcement phase, its distribution in the matrix, and the state of the matrix–reinforcement interface are key parameters for the understanding of the material's behavior. It makes sense to use the simplest and most widely available techniques first. Light optical microscopy (LOM) and scanning electron microscopy (SEM) provide a wealth of information about a material. However, in order to relate the observations to properties, a quantitative description of the microstructure is needed, and hence quantitative image analysis techniques are required. A brief introduction to this subject is given here, as it is not provided elsewhere in these volumes.

12.2 Sample Preparation

12.2.1 Purpose

The purpose of sample preparation has been described well by Clyne and Withers [1]: 'The aim of metallographic preparation

is to prepare a surface such that one or more of the following features can be clearly delineated without unnecessary distortion or the introduction of new features not originally present:

- distributions of reinforcement size, shape, orientation and location;
- matrix grain structure, presence of precipitates, etc.;
- information relating to crystallographic texture and distribution of plastic strain.'

This task is often difficult, due to the large difference in properties between the matrix and the reinforcement material. Artifacts, such as matrix smearing, reinforcement cracking, fiber pull-out, or interface damage, can easily be introduced into the material if proper care is not taken.

Detailed instructions for preparation are hard to find, except for the more intensively studied systems. Several authors, however, give general instructions on composite material sample preparation [1, 2].

12.2.2 Sectioning

The aim of sectioning is to produce a specimen with dimensions that allow further handling, near or exactly at the plane of interest. This stage of the preparation process involves the highest risk of damaging the specimen.

In hard reinforcement–ductile matrix systems, the matrix microstructure is vulnerable to change by excessive deformation introduced by the cutting operation. For thermoplastics with a low melting point (e.g. poly(ethylene) or poly(propylene)), excessive heat generation during cutting can melt the matrix, thus completely altering the microstructure. The use of abundant fluid cooling is advised, both for cooling and for rapid removal of cutting debris from the cutting slit. Burring and reinforcement pull-out at the edges can be minimized by mounting the specimen in resin prior to cutting (see Sec. 12.2.3 of this Chapter).

CMCs are normally cut with low-concentration diamond wafering blades, because they deliver an increased load per abrasive particle at the specimen–blade interface, which helps overcome the ceramic's hardness [3]. Blades having large abrasive particles and blades with fine abrasive particles are available. For brittle, surface-sensitive materials that can be damaged by coarse diamond particles, blades with finer particles are recommended.

For most MMCs the least deformation is introduced by using a low-speed diamond wire saw, or spark erosion. However, these techniques are slow. Cut-off wheels (diamond or ceramic impregnated) are quicker, but cause more matrix material to be deformed. Consequently, the first grinding step should be prolonged to remove the deformed layer.

For some PMCs, water-jet cutting or laser cutting procedures are the only means to produce a clean cut and a representative cross-section.

12.2.3 Mounting

Mounting has several purposes. One is to facilitate the further handling of the specimen during either manual or automatic preparation, by giving it a suitable

size and shape. The mounting also supports fragile specimens and ensures good edge retention. The abrasive properties of the mounting resin should be similar to those of the composite, otherwise edges will be damaged by the uneven removal of material. With temperature-sensitive materials, one must take care that the heat generated by the setting of the resin does not cause damage.

Specimens that are porous and cracked (e.g. not fully dense CMCs, or fiber composites that have not been completely infiltrated) should be impregnated with a low-viscosity resin to fill the pores. Otherwise, grinding and polishing debris may accumulate in the pores, extensive reinforcement breaking and pull-out may occur in the neighborhood, and bleeding out of etchant from the pores may reduce the quality of observation after sample preparation. The use of vacuum impregnation is highly advisable. In this case, resin is poured under vacuum into the mold containing the specimen, resulting in better infiltration of the open areas by the resin.

12.2.4 Grinding

The aim of grinding is to remove the damaged layer and the surface relief caused by sectioning, and to obtain a planar surface. Grinding is a multistep procedure; the optimum number of grinding steps and the abrasives used depend on the material. The most common abrasives are diamond and silicon carbide. The latter is usually bound to a paper substrate; abrasive particle sizes range from approximately 100 to 4000 mesh. Diamond is used either mounted in metallic or polymer substrate (grinding wheels) or suspended in a liquid on a rotating disk (lapping). Diamond powder sizes range from 10 to 100 μm.

If the cutting procedure does not introduce much damage in the sample, one step using a fine grinding medium (20 μm diamond or 600 mesh SiC) may be sufficient. Otherwise, several steps are necessary, progressing from coarser (sometimes as coarse as 100 μm diamond or 180 mesh SiC) to finer media (10 μm diamond or 4000 mesh SiC). If the matrix is ductile or heat sensitive, a high water flow is recommended for cooling and the removal of debris. Between steps, the sample must be cleaned thoroughly to remove all diamond or SiC that might still adhere to the surface.

When grinding and polishing MMCs, the use of automated machinery is recommended, because of the higher reproducibility of results, and the high and even pressure that can be applied. Grinding times should be kept to a minimum (at most a few minutes), and the sample should be lifted frequently from the grinding surface to avoid excessive heating and debris pile-up. Most authors recommend the use of high pressure. In the case of composites containing SiC particulate reinforcement, grinding with SiC papers can introduce foreign SiC particles from the grinding paper into the sample.

For PMCs, the grinding pressure should be minimal to avoid distortion of the matrix and pull-out of the fibers.

For composites made of hard, brittle materials, such as CMCs, it is preferable to achieve gross material removal during sectioning rather than by grinding with coarse abrasives. The properties of the

composite's most brittle constituent dictate the selection of the abrasive size. CMCs are ground using diamond abrasives (15–45 μm), either fixed (diamond wheels) or free (lapping substance). One way of grinding CMCs is to use 40 and then 20 μm nickel-bound diamond wheels, followed by 30 and then 10 μm plastic-bound diamond wheels. The change from nickel-bound to plastic-bound diamond wheels requires a step back in the coarseness of the abrasive particles because the nickel-bound wheels are more aggressive than the plastic-bound ones, thus creating more damage in the composite.

Another technique that is suitable for grinding CMCs is power lapping [3], where a load is applied on the specimen during lapping. This introduces less damage into the material than grinding, and appears to be the best method to maintain the specimen's flatness, especially when the constituent phases are of very different hardness.

12.2.5 Polishing

The aim of polishing is to obtain an optically flat surface, which presents the true microstructure of the material to the observer.

Mechanical polishing is done using cloths impregnated with fine diamond suspensions. Napless or low nap cloths are preferred, as these tend to give minimal surface relief during polishing. They are, however, more prone to contamination by polishing debris, so an adequate lubricant flow is necessary if the matrix is ductile, and cloths should be cleaned well after use. Sometimes, a diamond slurry is used on a ground-glass plate substrate for coarse polishing. Polishing is done in several steps, using decreasing diamond sizes (starting sizes range from 6 to 9 μm; final sizes range from 0.5 to 1 μm). A final step can consist of a short polish with colloidal silica.

For MMCs, the polishing-wheel velocity is usually kept low, and high pressure is applied to the samples. Polishing times range from 5 to 1 min per step, the first steps usually being longer. As an alternative, and for PMCs, low pressure can be applied, and the polishing times made longer (10–45 min per step). Samples are thoroughly cleaned between polishing steps.

For CMCs, the highest material removal rate is obtained with mechano-chemical polishing [3] (fine diamond abrasive suspended in a chemically corroding fluid), using automated specimen preparation with high pressure on the composite. The increased applied load enhances the chemical activity of the polishing mixture due to frictional heating. Long polishing times tend to increase the phase relief within the CMCs and should be avoided. A typical two-step polishing procedure for CMCs is the use of a nylon cloth with a 3 μm diamond solution followed by the use of a high-nap cloth with OPS solution (0.1 μm silica suspension, pH $\simeq$ 8).

Anodic electropolishing can be used for aluminum matrix composites. Of course, it is necessary that good electrical contact can be made with the back of the sample.

12.2.6 Etching

The aim of etching is to enhance the visibility of microstructural features by bringing them into a chemical environment where the features one wants to

observe react differently from the rest of the material. In many cases, the material contrast between matrix and reinforcement is sufficient to make further etching unnecessary for the study of the reinforcement distribution. Etching, however, can reveal grain-structure details such as grain boundaries, the texture of phases, the state of crystallization, and the chemical segregation within a phase.

For MMCs, the classical etchants and etching techniques used for the unreinforced matrix metal can be applied. Problems arise if the matrix–reinforcement interface is attacked preferentially to the grains and grain boundaries; reinforcement dropout and highly uneven etching can result. If the material is slightly porous, etchant bleeding from the pores can severely deteriorate the sample quality. However, the latter problem becomes an advantage when one is interested in locating pores.

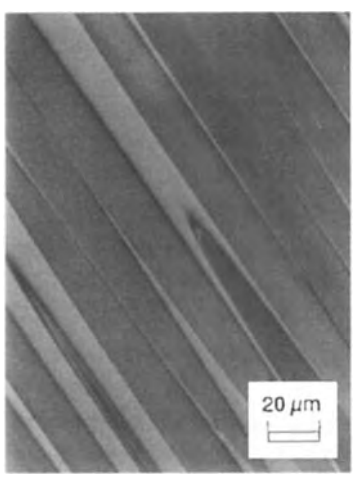


Figure 1. Glass–epoxy composite cross-section (LOM, ×500). The glass fibers were etched for 45 min using a solution of 10 g chromium oxide in 90 ml distilled water, resulting in an improved contrast between fiber and matrix.

For PMCs, etching is only used in special cases where the contrast is low. However, contrast between the matrix, the fibers, and the embedding resin is necessary, especially for quantitative image analysis. In the case of glass-fiber-reinforced epoxies and polyesters, different etching procedures have been described [4]. Figure 1 shows a glass epoxy composite etched in a chromic acid solution.

Sometimes the residual glass in the matrix of CMCs is etched with HF solutions in order to reveal the development of crystalline phases.

12.3 Quantitative Microstructural Analysis

12.3.1 Introduction

Quantitative microstructural analysis (in metallurgy also called quantitative metallography) is based mainly on classical geometrical and statistical concepts, and has a history of well over a century [5, 6]. In the last few decades, the concept of fractals has been introduced; this has found some application in describing and measuring the shape of microstructural constituents, a field which is still expanding rapidly. The advent of cheap and powerful computers has spawned the field of automated image analysis [7, 8], where the often tedious procedures are carried out by a computer. Automated image analysis has greatly facilitated the quantification of microstructures, but the specimen-preparation requirements are even more strict than for manual operations, because computers are often unable

to differentiate artifacts introduced by errors in grinding and polishing. As they become more affordable, techniques taken from medical imaging (tomography) may begin to find application in image analysis. The same applies to artificial intelligence systems, which not only quantify a microstructure but also interpret it. The importance of metallography for composite materials lies mainly in the measurement of the reinforcement distribution parameters, and of the state of the matrix–reinforcement interface.

Here, attention is given to some geometrical parameters that can be estimated from the study of plane sections through the material. Some material parameters, such as the amount of fiber fracture and fiber aspect ratio, are better determined by selective leaching of the matrix metal and studying the remaining 'naked' reinforcement. The distribution of chemical species throughout the material volume, which can, for example, be measured using electron probe microanalysis (EPMA), is not considered here.

12.3.2 Concepts of Quantitative Microstructural Analysis

There are five basic types of quantity of a microstructure that can be measured (Table 1). To each of these quantities can be added a subscript which indicates the reference for the measurement (Table 2, Fig. 2). When the measurement is related to the entire sample or when no ambiguity exists, no further subscripts are given; otherwise extra subscripts are added to indicate the meaning, for example, $(A_V)_{SiC}$ can be the surface/volume ratio of the SiC phase in a composite system.

A number of fundamental relations can be proven to exist between these quantities. The relations can be used to calculate microstructural properties that are difficult to measure from quantities that can more easily be measured:

$$V_V = A_A = L_L = P_P$$

$$S_V = (4/\pi)L_A = 2P_L$$

$$L_V = 2P_A$$

$$P_V = \tfrac{1}{2}L_V S_V = 2P_A P_L$$

These relations are general. That is, they apply in principle to all microstructures, whatever their geometry. It must be kept in mind, however, that they only hold exactly for 'true' geometrical properties. The values that are used in practice are obtained by measuring, and are thus subject to two types of error: random error and systematic error (or bias). The former is inherent to all sampling and measuring procedures, and can be reduced by increasing the number of samples or measurements. The latter is introduced by

Table 1. Types of quantity of a microstructure that can be measured

Type	Example	Notation
Volume	Volume of a phase	V
Area	Grain boundary area	A
Length	Perimeter of an inclusion	L
No. of entities	No. of inclusions	N
No. of points	No. of points of a grid located on inclusions	P

Table 2. Quantities related to microstructure

Relative quantity	Meaning	Example
V_V	Volume per unit volume	Volume fraction of a phase
A_V	Area per unit volume	Interface area per unit volume
L_V	Length per unit volume	Dislocation line length per unit volume
N_V	No. per unit volume[1]	No. of inclusions per unit volume
P_V	Points per unit volume[1]	
A_A	Area per unit area	Area fraction of a plane section occupied by a phase
L_A	Length per unit area	Particle perimeter per unit area
N_A	No. per unit area[1]	No. of cracks per unit area
P_A	Points per unit area[1]	No. of grain boundary triple points per unit area
L_L	Length per unit length	Fraction of the length of a line lying on a phase
N_L	No. per unit length[1]	No. of inclusions crossed by a line of unit length
P_L	Points per unit length[1]	No. of grain boundaries crossed by a line of unit length
P_P	Points fraction[1]	Fraction of points of a grid falling on a phase

[1]The difference between N and P quantities is not always clear. In general, P will be used when point-like features are counted, and N when the counted entity has a non-zero dimension (which is not considered relevant for the measurement). Also, P_V, P_A, and P_L generally concern point-like features of the microstructure, while P_P concerns points of some grid superposed on the microstructure.

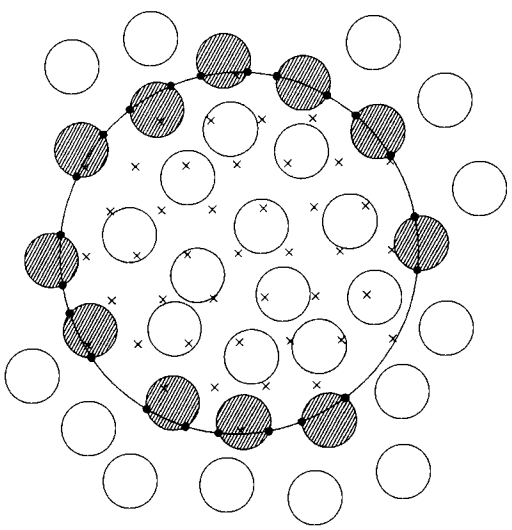


Figure 2. Different point and feature-counting measurements. Circles are the second-phase particles, the background is the matrix. The large circle and the crosses are a measuring line and grid projected onto the microstructure. P_L, the number of interfaces crossed per unit line length (number of black dots divided by the perimeter of the large circle); N_L, the number of inclusions intersected per unit line length (number of hatched particles divided by the perimeter of the large circle); P_P, the fraction of grid points that lie on particles (number of crosses inside particles divided by the total number of crosses).

incorrect sampling procedures, and cannot be reduced by repeated measuring. The above relations are valid for the expected value of the measured quantities, which is approximated by the mean of all unbiased measurement results.

The risk of measurement bias is smaller if the microstructure is 'random', that is if any section taken from the sample is equally well representative of the entire microstructure. Otherwise, the experimenter should carefully select those sections that are most representative of the quantities he wants to measure, and check the results for consistency with the assumptions that were made when choosing the section.

12.3.3 Sources of Error

12.3.3.1 Specimen Preparation Errors

Several possible causes of error have already been mentioned in the section on

specimen preparation:

- Matrix melting/deformation/cracking.
- Reinforcement deformation/cracking/pull-out.
- Interface cracking.
- Reinforcement hidden by smeared matrix material.
- Introduction of SiC from the grinding paper.

12.3.3.2 Sampling Errors

These errors are introduced when the section being studied is not representative of the material. A common cause of this kind of error is when the orientation distribution of microstructural constituents is not homogeneous. A typical example is a matrix with aligned fibers. The fiber volume fraction, for example, is more accurately measured in a transverse section. An experimenter must decide if the quantity of interest is best represented by a transverse or a longitudinal section.

Another cause is when the scale of certain features is larger than the measured area. If, for example, the edges of a composite are depleted of reinforcement, this will go unnoticed if only a few small areas in the middle of the sample are studied.

A third and very important cause is operator bias. An operator can bias results by his choice of measurement areas: if only sites 'that look nice' are selected, certain material characteristics are more or less unconsciously ignored. Results can also be biased by the operator's interpretation of certain unclear phenomena. If a measuring grid point falls on a particle edge, one operator may mark it as 'in' and another as 'out', thus influencing the result of the counting. If an inclusion is of a slightly different shade of color than most others, one operator may judge it to be a foreign inclusion, while another will count it as a piece of reinforcement. Such errors are best avoided by having several operators repeat the same measurement.

It is absolutely necessary that the greatest care is taken to avoid all biasing during sampling. Sampling regions and sectioning directions should be chosen considering all factual information and reasonable suppositions about the material at hand, and all results obtained should afterwards be checked against such suppositions. Moreover, as much true randomness as possible should be introduced into choices between equivalent sections and areas to be measured.

12.3.3.3 Errors Due to Image Analysis

Metallography by image analysis is sensitive to some kinds of error caused by the nature of the image processing hardware. The basic unit of an image in a computer is a pixel; a pixel has one color as its attribute. Pixel size is determined by the nature of the image-recording device and the magnification used to record the image. Any image information that is smaller than the pixel size is lost. Also, information on objects that are only a few pixels in size is very approximate; in particular, information on shape and perimeter is error prone. Therefore, to minimize this error, the objects studied should at least be several tens of pixels in size (Fig. 3).

It is generally the case that pixels associated with a phase do not all have the same color attribute, but a range of attributes. In order for the system to be able to

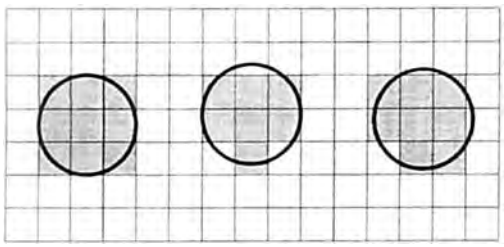


Figure 3. Very small objects cannot be represented accurately on a grid of pixels. The same small circle is represented by 7, 8, or 9 pixels, depending on its position relative to the pixel grid. Information on shape and perimeter is lost.

recognize the phases, the operator must decide which range of color values is to be associated with which phase. This is done by setting threshold values in the color spectrum. This process can easily lead to biasing error due to operator habits or optical illusions. This is particularly important in edge detection because, due to the analog nature of most recording devices, edges that are sharply delineated in the sample are seen as broader and fuzzier in the pixelized image. This fuzziness must be corrected for by applying sharpening filters, and setting appropriate threshold values. Significant errors can be introduced if this is not done correctly (Fig. 4).

Finally, if three phases are present in the system, one dark, one bright, and one of intermediate brightness, a halo effect can occur on the boundaries between the dark and the bright phases: the pixels of intermediate brightness on the fuzzy edge can be falsely interpreted as being of the third phase.

12.3.4 Reinforcement Volume Fraction

The reinforcement volume fraction $(V_V)_r$ is one of the most straightforward quantities to measure in a composite, and is one of the most important parameters governing the bulk properties of the material. There are three ways to determine the volume fraction, and these serve as generic examples for the three classes of measurements on plane sections.

One can determine the fraction of the measurement area that is occupied by the reinforcement $(A_A)_r$. This fraction is equal to the volume fraction or, more precisely, the area fraction is an unbiased estimate of the volume faction [9] (Fig. 5). Consider a cube of material with sides of length L, and a slice through this volume perpendicular to one side with infinitesimal thickness dx and area $A = L^2$. The reinforcement

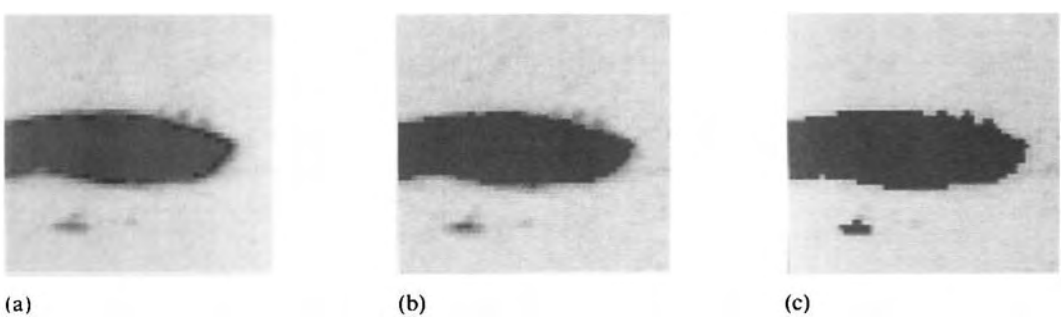


Figure 4. Edge error. The size of an object can vary significantly depending on how much of the fuzzy boundary around it is considered to be part of it.

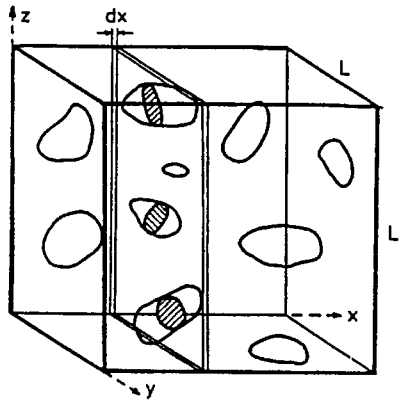


Figure 5. A cube with side L containing second-phase particles. dx is the thickness of an infinitesimally thin slice of the cube.

volume dV in the slice is equal to the area A_r of reinforcement in the slice, times the thickness of the slice:

$$dV = A_r\,dx$$

The total reinforcement volume V_r is the integral of dV over all slices through the cube:

$$V_r = \int_0^L dV = \int_0^L A_r\,dx$$

The average reinforcement area is:

$$\overline{A_r} = L^{-1}\int_0^L A_r\,dx$$

Then:

$$(V_V)_r = \frac{V_r}{L^3} = \frac{\overline{A_r}\cdot L}{L^3} = \frac{\overline{A_r}}{L^2} = (A_A)_r$$

where $(A_A)_r$ is the reinforcement area fraction. This is *areal analysis.*

When one draws an imaginary line segment across the plane section, and measures the fraction of the line length that is drawn over reinforcement areas, then this fraction is equal to the reinforcement volume fraction $(V_V)_r$. This is *lineal analysis.*

When an array of dots is drawn on the plane section, the number of dots that fall on the reinforcement divided by the total number of dots equals the reinforcement volume fraction. This is *point analysis.*

These calculations are, of course, only valid if the areas, lines, and dot grids used introduce no systematic error. This caveat is particularly important for fiber-reinforced composites; in this case the best method is to measure in a plane perpendicular to the fiber direction.

The statistical error of the measurement can be decreased by increasing the number of measurement areas, lines, or points in a section. When using arrays of parallel lines or dots, systematic error can be introduced if the reinforcement has a preferred orientation. This can be avoided by repeating the measurement with the array rotated over a small angle. It is good practice to prepare several sections of the sample for measurement in order to minimize the sampling error introduced by sectioning.

12.3.5 Homogeneity of the Phase Distribution

The failure of composites is greatly influenced by the homogeneity of the reinforcement distribution. Local concentrations of reinforcement can act as crack-generating sites; unreinforced regions are weak spots, where failure may begin.

Inhomogeneity can be described in different ways and can have various meanings. One meaning can be: 'the extent to which the particle/fiber arrangement deviates from a perfectly random arrangement' (or, conversely, a perfectly ordered lattice). This addresses inhomogeneity on the scale of a few reinforcement sizes, looking at how evenly spaced the

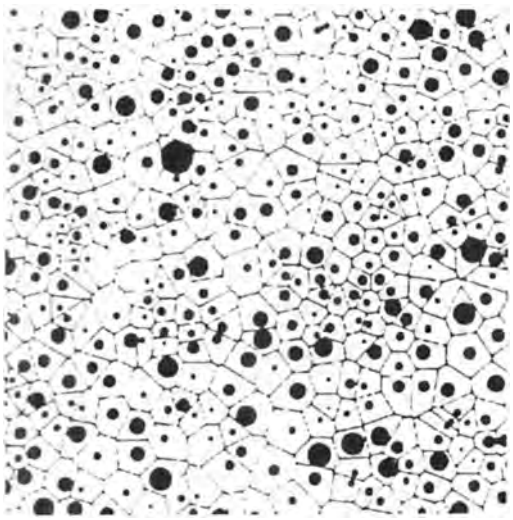


Figure 6. Dirichlet tessellation. Black circles represent reinforcement particles. The image is divided into cells. Each cell comprises the set of points that is closer to the particle enclosed within it than to other particles.

reinforcement is. Another can be: 'the extent to which the reinforcement volume fraction of different sample regions is different'. This addresses inhomogeneity on the scale of the sample, looking at whether there are regions that are less reinforced than others.

One way to address the small-scale inhomogeneity is to use Dirichlet tessellations. A Dirichlet tessellation is a network of polygons in which every polygon contains one particle or fiber, and all points in each polygon are closer to the enclosed particle than to any other particle (Fig. 6). From such a tessellation two parameters can be calculated: Q is the ratio of the observed nearest-neighbor distance to the one expected, and R is the ratio of the variance of the observed nearest-neighbor distance to the expected variance. The values of these parameters allow one to draw conclusions about the reinforcement distribution (Table 3) [10].

Table 3. Interpretation of the Q and R parameters obtained from Dirichlet tessellations

Q	R	Interpretation
<1	<1	Clustered sets
>1	<1	Short-range ordered sets
<1	>1	Sets of clusters with a superimposed background of ordered points

This approach becomes difficult if the sections through the reinforcement have very irregular shapes, as then different possible definitions of the distance from a particle can give rise to different tessellations and different values for Q and R.

A possible measure of large-scale homogeneity h is given by

$$h = \frac{1 - s_f}{f}$$

where f is the nominal volume fraction of reinforcement and s_f is the standard deviation on f [11]. Measuring h involves measuring the local reinforcement fraction over the sample (see Sec. 12.3.2 of this Chapter).

An alternative approach was developed for measuring the homogeneity of the fiber distribution in MMCs prepared by pressure infiltration. In this technique, a preform consisting of stacked fiber bundles, each containing hundreds of filaments, is infiltrated by liquid metal. As the metal tends to separate the bundles rather than to infiltrate between the individual filaments, a structure is formed of regions with a high and rather homogeneous reinforcement fraction, separated by regions containing no reinforcement (Fig. 7a). A measure of the homogeneity of such a

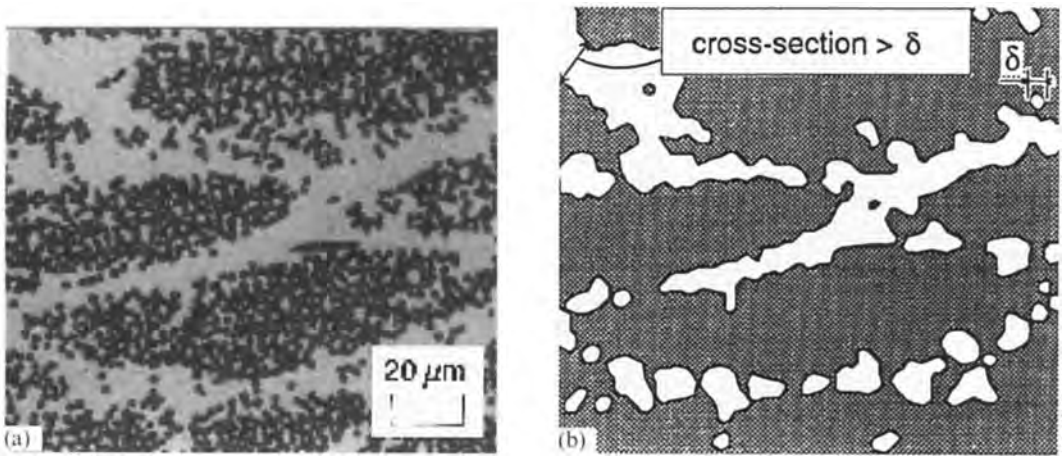


Figure 7. Determination of the unreinforced area fraction U, the ratio of the white area over the total section area, with a resolution of δ. (a) Micrograph of a transverse section through a 6061 aluminum composite reinforced with 50 vol.% SiC fibers. The material was prepared by infiltrating prestacked fiber bundles with liquid aluminum; the aluminum tends to separate the bundles (LOM; no etching). (b) Drawing derived from the micrograph in (a) by image processing. The white areas are regions with a minimum dimension of δ, that is areas which contain no fibers.

fiber distribution is the unreinforced area fraction U in a section. To determine U, a circle of diameter δ is drawn about every point in the section, and the area A of the union of all circles that contain no reinforcement is determined (the white area in Fig. 7b). U is the ratio of A to the total area of the section.

As δ decreases, the fineness of the measurement will increase and U will become larger. This allows one to tune the measurement to the scale of homogeneity one wishes to measure. From the variation of U with δ, information on the fractal geometry of the unreinforced areas can be deduced. δ_{min}, the smallest meaningful value of δ, is the nearest-neighbor spacing of the reinforcement arranged on a perfect lattice [12]. It can be shown [12] that if $\delta = \delta_{min}$, then

$$h = 1 - \sqrt{\frac{1}{1-U} - 1}$$

12.4 Structure of the Constituent Phases

12.4.1 Matrix

For PMCs in the form of a three-dimensional, woven, sandwich fabric, a lightweight foam is used as a core material. Although the foam properties are often neglected in calculations, they do have an influence on the mechanical performance of the sandwich. These properties are determined by the geometry of the foam cells. In order to model and predict the mechanical performance of the foam, it is essential to quantify all the cell dimensions [13]. The fabric is therefore cut at cryogenic temperatures in order to minimize distortion of the cell struts and membranes. The cell dimensions can then be measured from SEM images of the cross-sections (Fig. 8a). At high magnification

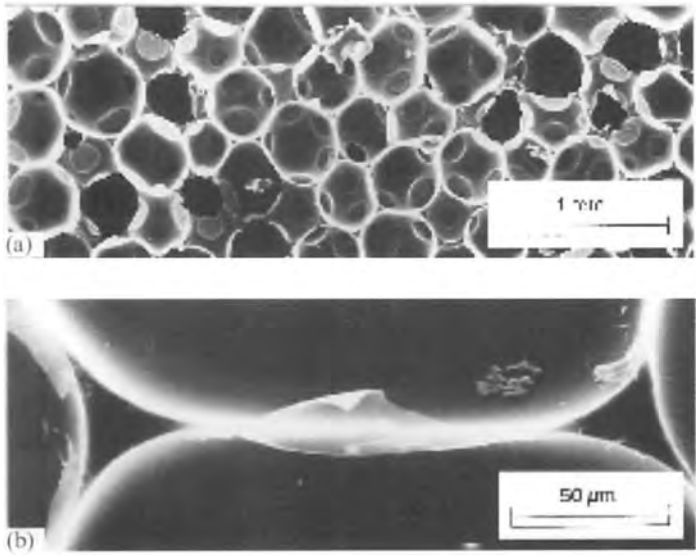


Figure 8. (a) SEM image (×28) of a cross-section obtained by cryogenic cutting of polyurethane foam. (b) SEM image (×500) of a cell membrane and two struts of polyurethane foam.

even the thickness of the membranes can be measured. Using an image like the one in Fig. 8b the thickness of struts and membranes can be measured, although the margin of error is very large. No alternative measurement techniques exist at present.

If polyurethane foam is injected into the core of a composite, the flow of the foam during expansion is strongly influenced (disturbed) by the presence of the pile fibers. From SEM images such as the one shown in Fig. 9, the size of the zone affected and the cell dimensions in the vicinity of the pile fibers can be quantified. This local geometry has a strong influence on the support of the pile fibers during mechanical loading of the sandwich beam.

Glasses and glass-ceramics are often used as matrix in CMCs, because their viscous flow properties at high temperatures allow for relatively easy densification

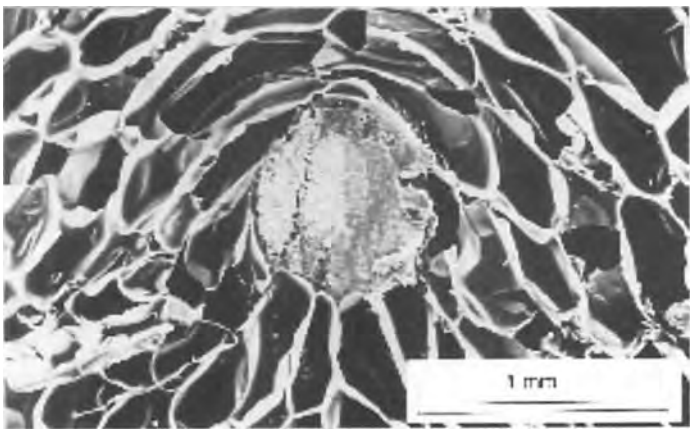


Figure 9. SEM image (×41) of the core of a three-dimensional sandwich fabric panel, containing polyurethane foam. The glass pile fiber in the fabric disturbs the flow of the polyurethane, resulting in a wide variation in the foam morphology.

Figure 10. Microstructure of a BMAS glass-ceramic, heat treated at 1070°C for several hours. Quantitative compositional EPMA indicated the presence of fine, interpenetrating, white celsian and black cordierite crystals, in combination with isolated barium osumilite crystals. A small amount of residual glass surrounds the crystalline phases. Back-scattered electron detector image (×2000).

of the composite. Glasses have an amorphous structure, although they usually show short-range order [14, 15]. Glass-ceramics can first be densified in the glassy state and then subsequently turned into a refractory polycrystalline material by means of an appropriate heat treatment [14–16]. Their properties are very closely related to their microstructure and, more specifically, to: (1) the amount and type of phases present, including porosity; (2) the characteristics of each phase, such as size, shape, and orientation; (3) the elastic modulus mismatch and the thermal expansion coefficient mismatch of the different phases; and (4) the strength of the interface between the phases present [14, 15]. The microstructure of glass-ceramics depends strongly on the processing conditions, such as temperature, holding time, and pressure; several microscopy techniques can be used to illustrate this dependence. The complexity of the microstructure of a $BaO–MgO–Al_2O_3–SiO_2$ (BMAS) glass-ceramic with no reinforcement is illustrated in Fig. 10.

12.4.2 Fibers

In composites reinforced with textiles, which are often used in PMCs and in some CMCs, the geometry of the textile is of primary importance for the mechanical properties and the (de)formability of the composite structure. In models that link this textile geometry to the mechanical properties (stiffness and strength) of the composite, a detailed description of the fabric is necessary. This description is obtained using stereo-microscopy. For a woven fabric, the geometric parameters can be obtained easily. However, for a complex knitted fabric such as the one shown in Fig. 11, a two-dimensional description of the fabric is obtained. The third dimension is obtained by taking into account the relative positions and dimensions of the fiber bundles. Subsequently, the fiber bundles are split up into small, straight segments, each having a specific orientation. These orientations, translated into an orientation distribution function, are used to calculate the elastic properties of the knitted fabric composite [17]. The same technique is used to characterize the deformation behavior (yarn slip, loop

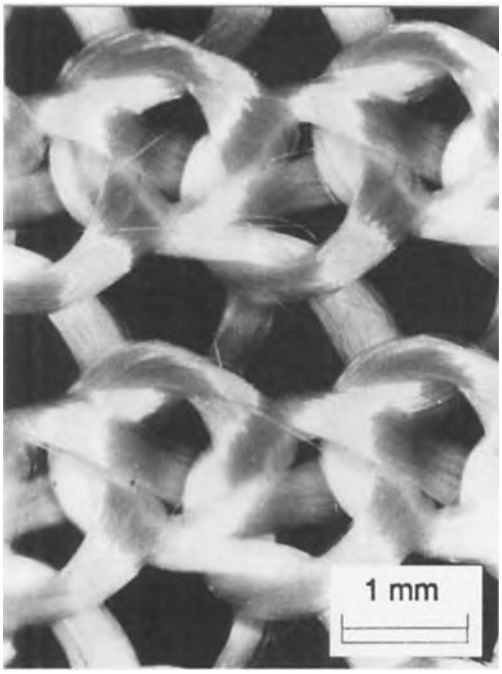


Figure 11. Stereomicrograph (×17) taken from a warp-knitted glass fabric, showing the loop structure.

closing and jamming, and yarn straightening). This is of utmost importance for the application of knitted fabrics in composites, as their main advantage is their high deformability which can be exploited to form complex-shaped parts.

The strength of the ceramic fibers used in CMCs is continuously degraded during processing of the composite, and thus the remaining strength is an important parameter when modeling the properties of the composite. The strength of the fibers in pristine condition or in the fabricated composite can be deduced from fracture mirrors after the composite's failure. A fracture mirror is a smooth, mirror-like region surrounding the fracture-initiating defect in the fiber [18, 19]. The size of the mirror radius a_m can be related to the fiber tensile strength S according to the equation [19, 20]:

$$S a_m^{\frac{1}{2}} = \text{constant}$$

By measuring the fracture mirrors on many pull-out fibers on the fracture surface of a failed composite, it is possible to construct a cumulative probability distribution $P(S)$, which represents to some extent the in situ fiber strength distribution [21]. The fracture surface of a fiber found within a failed Tyranno fiber/BMAS glass-ceramic matrix composite is shown in Fig. 12. Tyranno is a commercial silicon carbide (SiC) fiber containing 1.5–4 wt.% titanium.

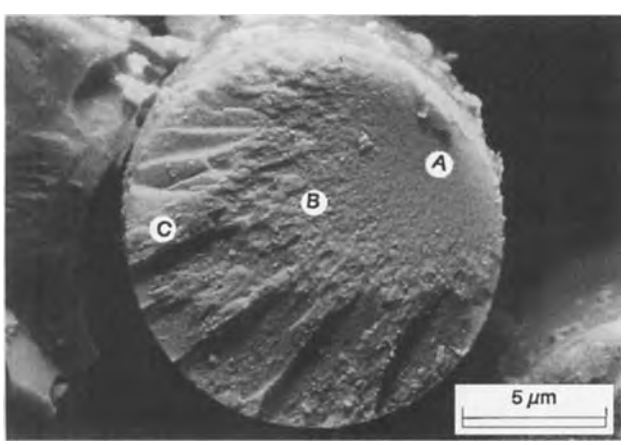


Figure 12. High resolution scanning electron micrograph (HRSEM) of the fracture surface of a Tyranno fiber. Three regions of decreasing smoothness are observed around the fracture-originating flaw: (A) the 'mirror' region; (B) the 'mist' region; (C) the 'hackle' region. The displayed fiber was found on the fracture surface of a Tyranno fiber/BMAS glass-ceramic matrix composite. Back-scattered electron detector image (×5000).

12.5 Structure of Composites

12.5.1 Interaction Between Matrix and Fibers

The crystallinity of thermoplastic polymers and composites is studied using thin slices of the material. The microstructure is revealed by the use of transmitted polarized light. Figure 13a shows a pure polymer: spherulites nucleate at random sites and grow until they meet. The presence of fibers in the molten thermoplastic changes this crystallization behavior: as shown in Fig. 13b, the nucleation points are no longer random, but nucleation starts from the fiber surface. Only in areas of low fiber density will spherulites nucleate in the matrix. The resulting microstructure shows much smaller spherulites. The size of the spherulites has a strong influence on the mechanical properties of the matrix [22]. This is one of the reasons why the mechanical properties, such as the fracture toughness, of the thermoplastic matrix cannot be transferred directly to the composite properties.

In some textile composites, the fabric geometry is stable only after production of the composite. A typical example is the three-dimensional woven sandwich fabric. During production, two skins interconnected by fiber bundles in the third dimension ('pile bundles') are pulled apart to produce a sandwich. The stretching forces applied determine the shape of the pile bundles. This shape (Fig. 14) has an important influence on the compression and shear strength of the core, as well as the stiffness of the core. In order to model these properties, an accurate description of the pile-bundle shape is required. This is obtained by cutting the composite in two orthogonal directions (warp and weft), and examining the shape in both planes

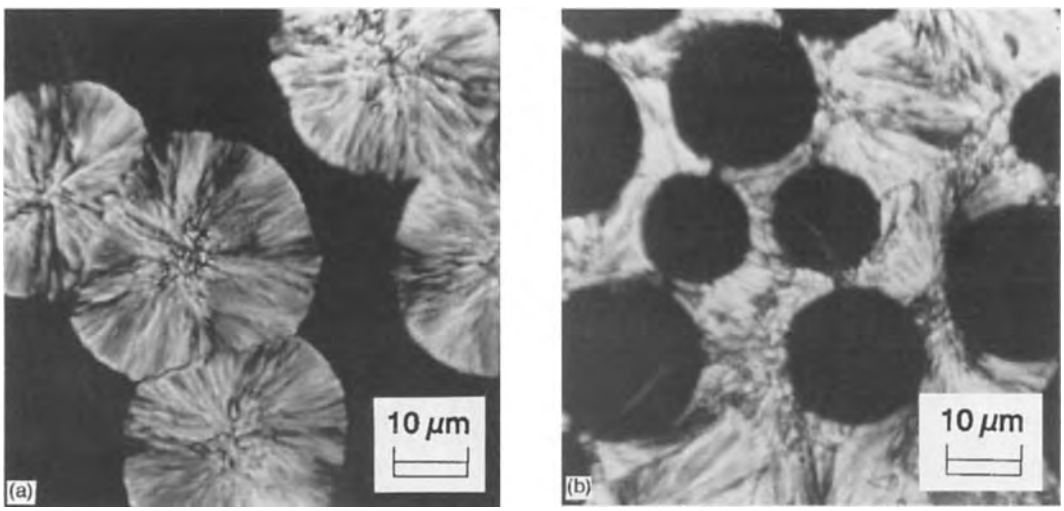


Figure 13. (a) Transmission light microscopy (×1000, using polarized light) reveals the presence of crystalline spherulites in a thermoplastic PAA (Polyarylamide) material which was quenched to freeze at an intermediate stage of the crystallization process. (b) The same technique was used to illustrate the microstructure of a composite with a crystalline thermoplastic PAA matrix. The spherulites nucleate from the surface of the glass fibers (×1000).

Figure 14. SEM image (×10) of the core of a three-dimensional woven sandwich fabric composite. The shape of the connecting pile fibers depends on the production process, and can be quantified using microscopy techniques.

by means of stereo-microscopy. The data from both images are subsequently combined to produce a three-dimensional representation which is used for automatic mesh generation in a finite-element model pre-processor.

SEM techniques are often used to depict microstructural features of CMCs, such as the homogeneity of distribution of the reinforcing phase, the degree of porosity, the development of matrix microcracking, or the structure of the matrix, reinforcement, and interface

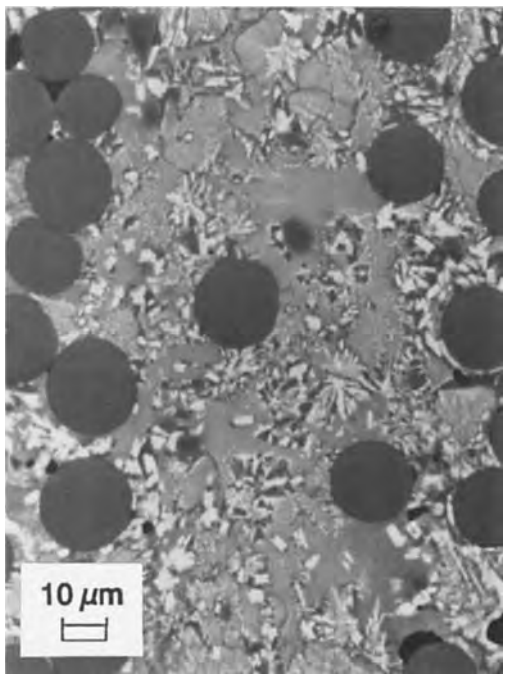


Figure 15. EPMA micrograph (back-scattered electron detector image, ×600) of a Nicalon fiber/BMAS glass-ceramic matrix composite. The matrix consists of barium osumilite, celsian, and cordierite crystals surrounded by residual glass. Fiber-distribution inhomogeneities and matrix porosity are shown.

between them. The microstructure of a Nicalon (commercial silicon carbide) fiber/BMAS glass-ceramic matrix composite is shown in Fig. 15.

12.5.2 Characterization of the Interface

The fragmentation test is commonly used to characterize the interfacial shear strength in PMCs. In this test, a single fiber is embedded in a transparent matrix with a high strain to failure. During loading, the fiber breaks into small fragments. The most commonly used data-reduction

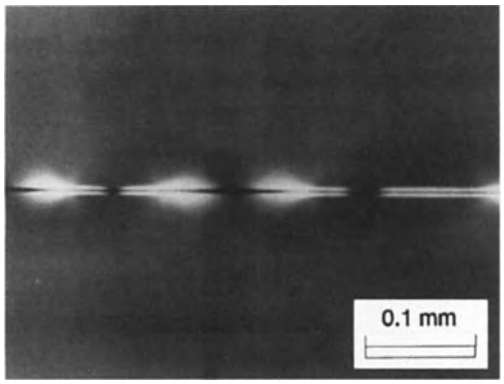


Figure 16. Transmission light microscopy with polarized light (×200) reveals the fiber breaks (dark spots) and the length of the fiber–matrix interface debonding (narrow white fringes close to the fiber) during a single-fiber fragmentation test on a carbon-epoxy composite specimen.

schemes use only the fiber fragment length to characterize the interfacial shear strength. Studies [23] have shown that these data-reduction schemes are not correct. Moreover, it has been demonstrated that, by using transmission microscopy with polarized light, not only the fragment length, but also a debonding length, can be determined. In Fig. 16, the debonded area is revealed by narrow white fringes close to the fiber. The measurement of the debonding length allows a more detailed analysis of the test data, so that a more accurate value of the interfacial shear strength and the interfacial friction stress can be calculated [24].

12.6 Damage Mechanisms

12.6.1 Introduction

Modern engineering materials are required to sustain a reasonable level of damage before complete failure. Sudden brittle fracture without prior warning is not acceptable. This is particularly a problem for CMCs. For the materials engineer it is important to study the microstructural mechanisms by which a composite is able to sustain damage and by which it will ultimately fail. Using the results of such studies, damage tolerance can be designed into the composite.

Some examples of such studies are given below. Microscopy is usually essential to identify the particular damage modes that occur in the various types of composites.

12.6.2 Delamination of Fibrous Composites

PMC and CMC composites are often fabricated from laminae or plies. Each lamina is composed of a single layer of fibers, which have been impregnated with matrix material. Laminae are stacked together with the fibers in the desired relative orientation, and the stack is heat treated so that the matrix forms into a continuous phase. The volume fraction of reinforcement is usually lower between the laminae than within each ply. Hence, the interlaminar volume is often the source of damage in the composite, and delaminations are often an important damage mode in PMCs and CMCs.

In laminated plates, delaminations are created by out-of-plane loads (often due to impact by foreign objects). They are invisible and can grow very quickly during subsequent loading of the part, resulting in stiffness degradation and premature failure of the composite part. This is particularly critical in the very brittle phenolic

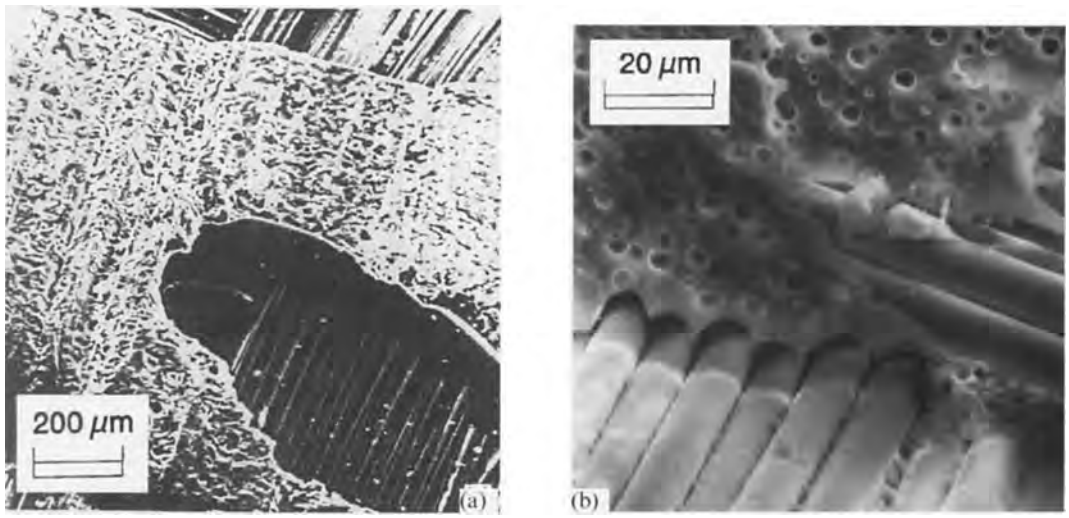


Figure 17. (a) SEM image (×50) of a delamination surface on a woven-glass fabric-reinforced phenolic resin. The delamination jumps from one fabric layer to another. The resin fragments clearly show the imprints of the glass fibers. (b) SEM image (×600) of the same delamination surface, showing that the delamination follows the fiber/matrix interface exactly (although a very thin matrix layer might still be present on the fiber surface).

resin composites [25]. Figure 17a shows a SEM micrograph of a delamination fracture surface of a woven fabric composite. The picture clearly reveals the fabric structure and a block of matrix, while imprints of the next fabric layer are visible in the matrix. The same figure shows that a crack jumps from one layer to the other (this often occurs due to the presence of large voids) and the crack path is near the fiber interface. Micrographs at higher magnifications (Fig. 17b) verify the presence of interface cracks and debonding.

Ply delamination is an important failure mode that occurs in laminated CMCs [15]. Figure 18 presents a central delamination observed in a failed silicon carbide fiber/borosilicate glass matrix composite.

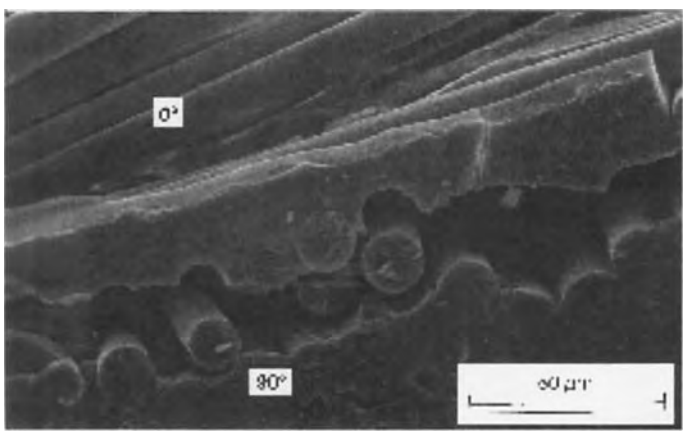


Figure 18. SEM image (secondary electron detector image, ×655) showing the central delamination in a cross-plied composite, consisting of Nicalon (silicon carbide) fibers and a Pyrex (borosilicate) glass matrix. This specimen was tested in three-point bending and the cracks propagated with preferential ease in the glass-rich area between the 0° and 90° plies.

Delamination cracks in cross-plied composites, such as the one shown in Fig. 18, propagate preferentially in the glass-rich layers between the 0° and 90° plies.

12.6.3 Interfacial Failure

The interface between the constituent phases can also be a weak area in a composite. This may actually be desirable in CMCs, but in the other types of composite it is usually not.

The development of damage in carbon-fiber-reinforced PMCs is also influenced by the interface properties. Carbon fibers are treated electrochemically to improve the bonding with the matrix. As a result, the fracture path changes [26], as shown in Fig. 19. In a composite containing untreated fibers, the fracture surface is directly at the fiber surface. Figure 19a shows a clean carbon fiber where the striations are still visible. Increasing the current density during electrochemical treatment results in a shift of the fracture path. In Fig. 19b, a thin resin layer (in this case with an unusual morphology, because the resin is a dual-phase system) is present on the fiber surface. A further increase in the treatment intensity results in a crack which runs through the matrix and the fibers are covered with a large amount of resin. The changes in the crack path have an important influence on the resulting mechanical properties of the composite.

The major drawback of monolithic ceramics for structural applications is their brittleness or, more accurately, their low fracture toughness, which leads to sudden catastrophic failure. A significant enhancement in the toughness of monolithic ceramics can be achieved via CMCs. This is due to the incorporation of a reinforcement (e.g. fibers or whiskers)

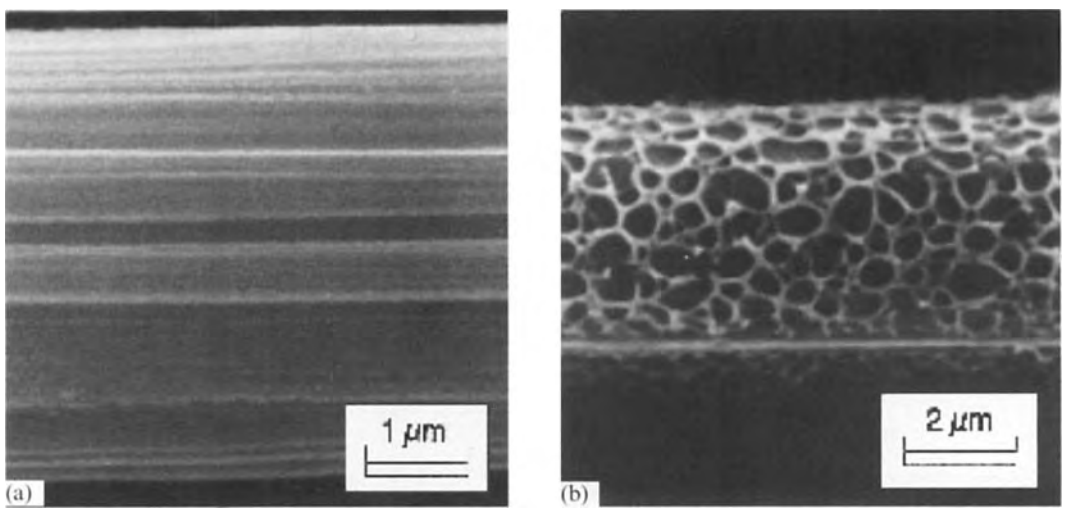


Figure 19. (a) SEM image (×11 400) of a carbon fiber in a fracture surface of a composite containing untreated fibers. The fracture surface has followed the fiber-matrix interface, demonstrating the low interface strength (C-Ep composite). (b) SEM image (×5950) of a carbon fiber in a fracture surface of a composite containing intermediately treated fibers. The low level of the oxidating surface treatment of the fiber creates an interphase between the fiber and the bulk matrix. The fracture surface follows the interphase layer (C-Ep composite).

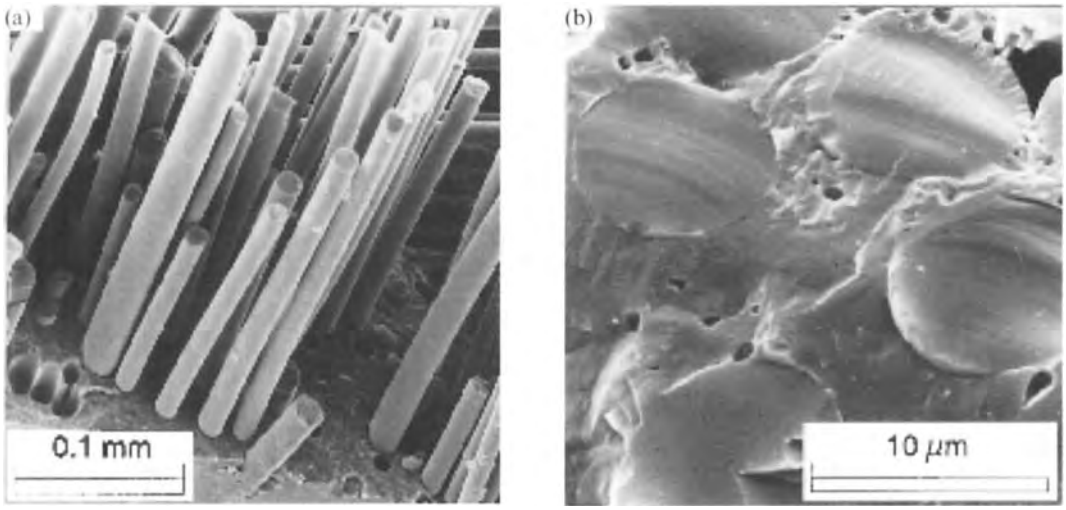


Figure 20. (a) SEM image (secondary electron detector image, ×186) of the fracture surface of a commercial Tyranno (silicon carbide with titanium nitride) fiber/BMAS (barium magnesium aluminosilicate) glass-ceramic matrix composite, the toughness of which is due to the large amount of fiber pull-out. (b) SEM image (secondary electron detector image, ×2620) of the fracture surface of the same composite after a heat treatment (900°C, 16 h), which resulted in degradation of the graphitic fiber-matrix interface. The failure changed from tough to brittle, with no fiber pull-out.

which may introduce into the ceramic matrix fracture energy-dissipating mechanisms, such as fiber-matrix interface debonding, crack deflection, fiber bridging, and fiber pull-out [15, 20, 27, 28]. Whether or not such toughness-enhancing mechanisms will in fact occur depends on the strength of the matrix-reinforcement interface. If the interface bonding between the fibers and the matrix is strong, an oncoming crack will not be influenced by it and this will lead to brittle failure of the composite. On the other hand, weak fiber-matrix bonding can serve as a crack arrestor, increasing substantially the fracture toughness of the composite [27, 28]. The properties of these interfaces need to be controlled closely during the production of CMCs; they may also evolve during service of the material.

The tough behavior of a commercial silicon carbide fiber/BMAS glass–ceramic matrix composite shown in Fig. 20a is characteristic of a class of composites that have a graphitic (C) fiber-matrix interface [27–30]. This interface is formed during hot-pressing of the composite according to the following oxidation reaction [29, 30]:

$$SiC(s) + O_2(g) \leftrightarrow SiO_2(s) + C(s)$$

and is responsible for the increase in the composite's fracture toughness.

Oxidative removal of the carbon interfacial layer followed by oxidation of the fiber surface and strong bonding between the fiber and matrix take place at temperatures above 500°C in air [29–31]. As a result, toughness is strongly reduced and this process severely limits the use of this class of composites at elevated temperatures. Figure 20b illustrates the brittle failure of the composite shown in Fig. 20a after it has been heat treated in air

and the graphitic fiber-matrix interface has been removed.

12.6.4 Particle Fracture

When a MMC reinforced with brittle particles is put under strain, the microscopic deformation cannot be homogeneous as the particles cannot be deformed. The macroscopic deformation of the material must, therefore, be accommodated either by plastic flow of the matrix around the particles, by interface debonding, or by particle fracture. The mechanism depends on the material and the stress state. The latter two mechanisms result in the presence of voids in the material, which can act as nucleation sites for larger cracks which in turn cause total failure.

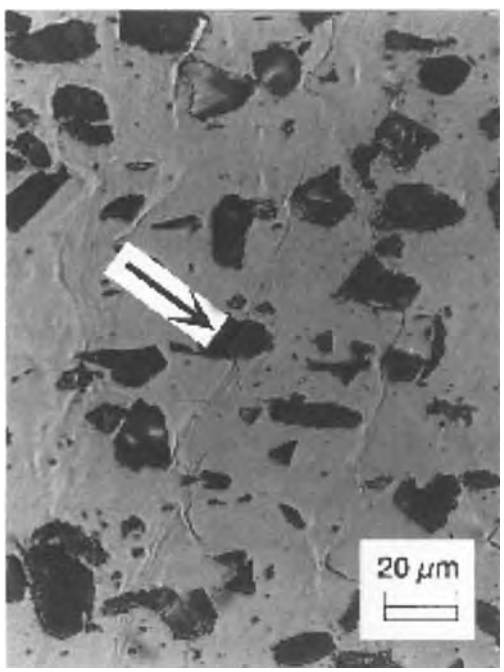


Figure 21. Particle fracture in 6061 aluminum containing 20 wt.% SiC particles. The material was first tempered to a T6 condition, after which the surface of the specimen was polished. Then 2.6% strain was applied, and the specimen was studied under the microscope as strained (LOM, using polarized light and Nomarski interference contrast). Arrow indicates fractured particles; fringes in the background material indicate localized strain.

It is not always clear from the study of fracture surfaces how the failure originated. To find this out, one must study the material during loading itself. One way to do this is by starting with a polished tensile specimen and studying it after different amounts of strain [32]. In this way damage features at the surface can be detected.

Figure 21 shows an Al–SiC composite studied under tensile loading. The particles crack to accommodate the plastic strain of the matrix, and these cracked particles act as nucleation sites for matrix cracks.

12.7 Concluding Remarks

In the present chapter the role of commonly available microscopy techniques such as LOM and SEM has been emphasized. Clearly, there is plenty of scope to use more advanced techniques, especially to characterize the interface between the constituent phases in composites. High resolution imaging and analytical techniques which have been used include transmission electron microscopy coupled with analytical techniques, such as energy dispersive X-ray spectroscopy, electron energy-loss spectroscopy, and Auger electron spectroscopy. For polymer-based composites X-ray photoelectron spectroscopy with high spatial resolution is appropriate because the technique also yields information on the type of bonding at

the interface. Secondary ion mass spectroscopy has also been used.

Internal stresses and strains are often an important characteristic of structural composites due to the differences in the thermal expansion behavior of the two phases. Indeed these play an important role in the selection of the constituent phases, particularly for MMCs and CMCs which are processed at high temperatures. Micro-Raman spectroscopy coupled with optical microscopy has found to be useful for characterizing composites with respect to internal stresses [33].

Finally, the importance of specimen-preparation techniques and quantitative image analysis in the field of composite materials cannot be overestimated.

Acknowledgements

The authors acknowledge support by the Belgian Ministry of Science Policy under IUAP contract No. 41. The research described was also supported in part by the Belgian Ministry of Education under GOA contract No. 92.96-01. K. Lambrinou acknowledges a fellowship from the EU under contract No. BRE2-CT93-3004. J. Ivens acknowledges a grant from the Flemish Institute for the Promotion of the Scientific–Technological Research in Industry (IWT). Finally, the authors would like to acknowledge S. De Bondt, B. Gommers, J.-F. Marsol, A. Niklas, E. Vanswijgenhoven, A. W. Van Vuure, and W. Wei for supplying some of the pictures.

12.8 References

[1] T. W. Clyne, P. J. Withers, *An Introduction to Metal Matrix Composites*, Cambridge University Press, Cambridge **1993**.
[2] L. L. Clements in *Metals Handbook*, 9th edn, Vol. 9, ASM, Ohio **1985**, pp. 587–593.
[3] D. C. Zippenian, S. N. Chanat, *Advanced Mat. Proc.* **8/93**, *144*(2), 15.
[4] B. J. G. De Aragão, *Pract. Met.* **1990**, *27*, 255.
[5] R. T. Dehoff, F. N. Rhines, *Quantitative Microscopy*, McGraw Hill, New York **1968**.
[6] E. E. Underwood, *Quantitative Stereology*, Addison Wesley, Reading, MA **1970**.
[7] H. E. Exner, H. P. Hougardy, *Quantitative Image Analysis of Microstructures*, DGM Informationsgesellschaft, Oberursel **1988**.
[8] M. Coster, J.-L. Chermant, *Précis d'Analyse d'Images*, CNRS, Paris **1985**.
[9] E. R. Weibel, *Lab. Invest.* **1963**, *12*, 131.
[10] H. Schwartz, H. E. Exner, *J. Microscp.* **1983**, *129*, 155.
[11] G. Ondracek, R. Renz, *Pract. Met.* **1989**, *26*, 368.
[12] S. De Bondt, Ph.D. thesis, Katholieke Universiteit, Leuven **1993**, pp. 30–33.
[13] P. De Neve, Masters thesis, Katholieke Universiteit, Leuven **1994**.
[14] W. D. Kingery, H. K. Bowen, D. R. Ulhmann in *Introduction to Ceramics*, 2nd edn (Eds.: E. Burke, B. Charmers, J. A. Krumhans), Wiley, New York **1976**.
[15] K. K. Chawla, *Ceramic Matrix Composites*, Chapman & Hall, London **1993**.
[16] S. H. Knickerbocker, A. H. Kumar, L. W. Herron, *Am. Ceram. Soc. Bull.* **1993**, *72*(1), 90.
[17] B. Gommers, I. Verpoest, P. Van Houtte, *Proc. EUROMEC 334: Textile Composites and Textile Structures*, Lyon, France **1995**.
[18] H. C. Chandan, R. D. Parker, D. Kalish in *Fractography of Glass* (Eds.: R. C. Bradt, R. E. Tressler), Plenum Press, New York **1994**, Chap. 5.
[19] W. A. Curtin, *J. Am. Ceram. Soc.* **1994**, *77*(4), 1075.
[20] A. G. Evans, F. W. Zok, *J. Mater. Sci.* **1994**, *29*, 3857.
[21] F. W. Zok, X. Chen, C. H. Weber, 'Tensile strength of SiC fibers', *J. Am. Soc.* **1995**, *78*(7), 1965.
[22] B. Goffaux, Ph.D. Thesis, Katholieke Universiteit, Leuven **1993**.
[23] M. Desaeger, M. Wevers, I. Verpoest in *Proc. ICCM-9*, Vol. 2 (Ed.: A. Miravete), Madrid, Spain **1993**, pp. 732–739.
[24] Th. Lacroix, R. Keunings, M. Desaeger, I. Verpoest, *J. Mater. Sci.* **1995**, *30*, 683.
[25] J. Ivens, A. W. Van Vuure, I. Verpoest, M. Fantino in *ECCM-7*, Woodhead Publ. Ltd., Cambridge **1996**, pp. 99–104.
[26] H. Albertsen, J. Ivens, P. Peters, M. Wevers, I. Verpoest, *Composite Sci. Technol.* **1995**, *54*, 133.
[27] J.-Y. Hsu, R. F. Speyer, *J. Mater. Sci.* **1992**, *27*, 381.
[28] R. F. Cooper, K. Chuyng, *J. Mater. Sci.* **1987**, *22*, 3148.

[29] K. P. Plucknett, S. Sutherland, A. M. Daniel, R. L. Cain, G. West, D. M. R. Taplin, M. H. Lewis, *J. Microsc.* **1995,** *177*(3), 251.
[30] M. H. Lewis, V. S. R. Murthy, *Composite Sci. Technol.* **1991,** *42*, 221.
[31] S. M. Bleay, V. D. Scott, *J. Mater. Sci.* **1992,** *27*, 825.
[32] A. Niklas, Ph.D. thesis, Katholieke Universiteit, Leuven **1993,** pp. 71–76.
[33] O. Van der Biest, T. Laoui, J. Vleugels, K. Sumanasiri, H. Mohrbacher, B. Blanpain, J. P. Celis in *Corrosion of Advanced Ceramics* (Ed.: K. G. Nickel), Kluwer Academic, Dordrecht **1994,** pp. 385–397.

Handbook of Microscopy
S. Amelinckx, D. van Dyck, J. van Landuyt, G. van Tendeloo

13 The Structure of Polymers and their Monomeric Analogs

13.1 Introduction

The proposition of writing a short chapter about polymers is a task like that of Sisyphus; since the problem cannot be solved, one may be persuaded of the absurdity of human endeavor. In addition to this, the result is almost certain to create everlasting animosity from respected colleagues whose excellent work has not been included. It is hoped that the reader will understand that only a small representative fraction of work can be included and that the choice is partly dictated by personal interest in and understanding of certain aspects of the field, but also in order to create some continuity within the paper. Only in this way will it be possible to communicate some of the fascinating problems related to this field, in which it is essential for chemists and physicists to work closely together, a field in which molecules can be manipulated in an apparently infinite number of ways to produce specific physical properties. For this reason also, this chapter emphasizes work in which new methods have been developed and innovative materials investigated.

The role of the electron microscopist is to find the equations which relate physical properties to molecular structure and superstructure and to find ways of making these structures visible. The qualitative assessment of an image, however good it is, is usually inadequate for an analysis of the physical properties. Thus quantitative analysis is frequently essential, not only to quantify the results, but also to find an adequate way of varying one parameter so that the data can be used to verify or disprove an equation.

The aim of this chapter is not to present a comprehensive overview of the whole topic (since this is impossible), but rather to highlight specific, characteristic developments and problems in the elucidation of polymer structure and defects, while keeping the mathematics to a minimum. The appropriate references are, however, given. The author hopes that some cross-correlation with other fields can be initiated when it is realized, with the help of interdisciplinary publications such as this book, that many similar problems occur in the solid-state physics of different materials and that science would benefit from an exchange of information and expertise.

The chapter is divided into four major sections: the role of electron microscopy in elucidating structure–property relationships in the major polymer classes; the

use of new microscopic methods to study polymers; electron crystallography of polymers and their monomeric analogs; and high resolution transmission electron microscopy (TEM).

A discussion of experimental high resolution TEM images of polymers and their monomeric analogs has been included because, strangely enough in view of the beam sensitivity and associated difficulties of TEM, some major advances in understanding the structure of polymers have recently been made using this technique. Images and diffraction patterns from polymers frequently do not achieve the quality of their monomeric analogs; polymer crystals tend to be smaller and less perfect. Therefore a study of the appropriate monomeric analogs delivers crucial information about the molecular packing and orientation. The literature has been chosen such that at least one standard textbook is included in each subsection and representative specialized literature is included which gives an extended reference list and an overview of the subject.

13.1.1 Investigation of Polymers by Electron Microscopy

The investigation of polymers by electron microscopy is driven by the desire to improve specific physical properties. The nature of polymeric chains, with their enormous range of molecular architectures, enables an apparently infinite variety of structures to be achieved.

Polymers or macromolecules are found naturally in the form of silk, rubber, wool, cotton, wood, etc., as well as being synthesized to form plastics, fibers, paints, coatings, and liquid crystals. They are giant molecules with molecular weights between 10^3 and 10^6, connected to one another by covalent bonds which may be synthesized to form linear chains, branched chains, or networks. The chains interact through van der Waals forces, while the individual atoms of the chains are linked by covalent bonds.

The type of atoms involved and their specific arrangement and length gives macromolecules their chemical distinctiveness and particular function. The complexities are far too great for a mechanistic deduction of properties from the detailed structure of the constituent molecules, and so a statistical approach has played a dominant role in the treatment of the composition, reactions, and physical properties of polymeric substances [1].

Polymers are particularly difficult to image in both TEM and scanning electron microscopy (SEM) because they are composed primarily of atoms with small atomic numbers and low densities, so that contrast is low. In addition, all aliphatic and some aromatic polymers are beam sensitive, so that they are destroyed almost before the image can be recorded.

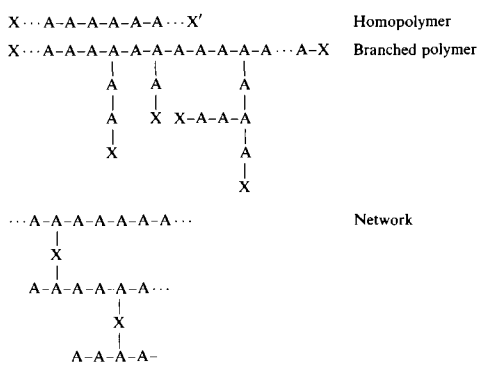
The interaction between the primary beam and the sample leads to many events which generate molecular fragments, secondary electrons, and X-rays. In addition, chemical changes such as cross-linking, degradation, and evolution of gases can occur at a rate dependent on the concentration of the respective species. A further disadvantage in the case of SEM is the nonconducting nature of polymers, which causes both charge and thermal build-up in the sample. If a polymer contains contaminants such as residual solvent, these may degas when placed in a vacuum and

cause polymerization reactions at the sample surface. This can give rise to a contaminating layer. Polymer microscopists have developed methods of preparation, recording, and analysis to cope with these problems, but they are generally specimen specific. For this reason, references have been chosen which clearly reveal the preparative procedures, but these cannot be treated in this article.

13.2 Structure–Property Relationships in the Major Polymer Classes

13.2.1 Homopolymers

Molecules composed of chemically and stereochemically identical units (apart from end-groups) are called 'homopolymers'; molecules with more than one kind of repeating unit are called 'copolymers'. If A represents the structurally repeating unit, and X,X′ are the end-groups, then geometries such as the following are possible:



The number of possible variations in chemical repeating units produces an enormous list of useful polymers. In addition to the many synthetic plastics, chain molecules such as polysaccharides, polypeptides, and nucleic acids can be organized into more complex macromolecular structures of great biological and biochemical importance. The effect of replacing chemical structural units on the physical properties is enormous. For example, if the hydrogen atoms in the linear homopolymer polyethylene are replaced by fluorine atoms, the melting temperature is increased by about 160°C and the coefficient of friction is decreased drastically. Alternatively, if aromatic rings are introduced into the chain backbone, the thermal and oxidative stability is enhanced and very rigid materials can be produced. Such polymers can replace metals in many applications.

The thermodynamics and statistical principles governing the behavior of polymers has been covered extensively by Flory [1]. Both the crystal structure and the superstructural arrangement of the crystalline entities are of paramount importance for the physical properties of these materials, and these are amenable to electron microscopic observation. Therefore homopolymers have been studied extensively by means of electron microscopy and electron diffraction.

13.2.1.1 Single Crystals

A large number of long-chain molecules can crystallize, depending on their molecular architecture and intermolecular forces. The crystals are generally much smaller than those of their monomeric analogs and are therefore inaccessible to light microscopy and X-ray structure

analysis, but are well suited to electron microscopy. The structure of homopolymer single crystals grown from solution and their relation to crystallization kinetics has been described in several books [2–4]. Some polymer single crystals have needle-shaped morphologies, but many polymers form flat or pyramid-shaped lamellar single crystals with a thickness which varies from about 100 to 500 Å and with lateral dimensions of several thousand angstrom or even tens of micrometers. The crystals vary in size, thickness, and shape depending on such factors as solvent, temperature, concentration, and rate of growth. Three typical examples are chosen to demonstrate the role that electron microscopy and diffraction has played in the elucidation of some fundamental problems related to these polymers. They represent the three major types of chain configuration: flexible chains, rigid chains, and helical chains.

Polyethylene

It has been shown by extensive electron microscopic studies [5–8] that polyethylene single crystals grown from dilute solution form hollow pyramids in solution which collapse when transferred to a substrate and form corrugations in the ⟨130⟩ directions in each of the four {110} sectors (Fig. 1b). Since electron microscopy has also played a decisive role in the interpretation of the fold surfaces in polymer single crystals, it is important to review this problem here.

The lateral extension of single crystals in the *ab* direction is generally in the micrometer range, whereas the crystallographic *c* direction, which is also the molecular direction, extends only to 100 Å, which is much less than the molecular length. It is clear, therefore, that the molecule has to fold at the crystal (lamellar) surface and to re-enter the crystal at a later stage during crystallization. The precise nature of this fold and the question of whether the molecule returns to this crystal in an adjacent or nonadjacent position was strongly debated in the literature for many years. Since the molecule emerging from the crystal surface is no longer confined to the closest packing induced by the crystal field, there is obviously a crowding problem at the lamellar surfaces where the molecule emerges. Density measurements clearly indicated that these surfaces consist of an amorphous overlayer [9–11]. This conclusion was supported by other evidence such as enthalpy of fusion and Raman spectroscopy measurements as well as other physical properties [12]. In contrast to this, infrared studies [13] and etching experiments were interpreted as indicating tight folding with strict adjacent re-entry [14].

Neutron scattering evidence from the small and intermediate scattering region of polyethylene single crystals indicated that some adjacent re-entry occurs at {200} growth faces, but that there is stem dilution along {110} growth planes [15, 16].

The role of electron microscopy in this important issue has been that the observation and analysis of sectorized crystals was interpreted for many years as clear evidence for tight folds with strictly adjacent re-entry. However, in a recent detailed comparative investigation of polyethylene and a series of nonfolded *n*-alkanes, it was shown that the relevant features of bright-field images related to the problem above are identical in unfolded *n*-alkanes and in polyethylene. These observations included the phenomenon of sectorization as well as

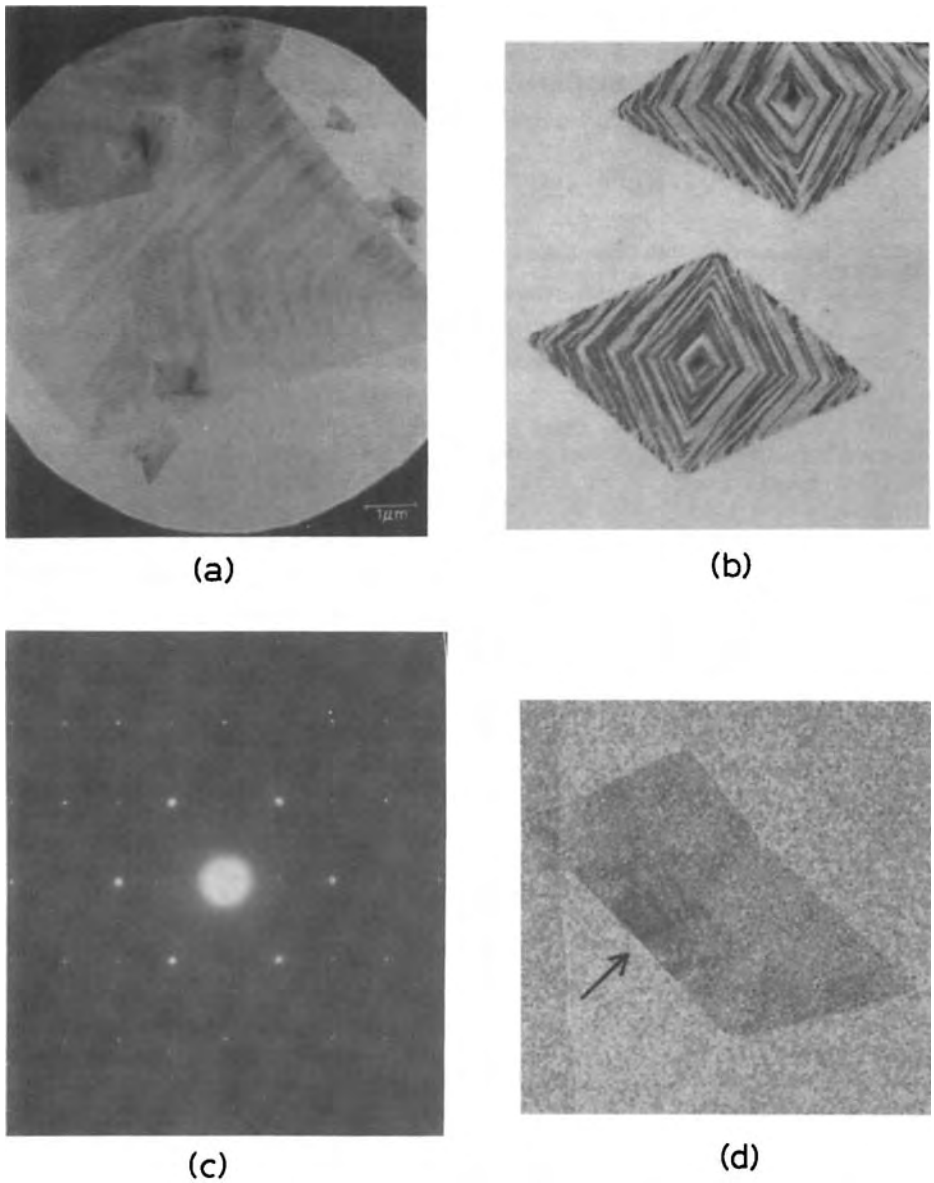


Figure 1. (a) Bright-field electron micrograph of a monolamellar crystal of paraffin n-$C_{82}H_{166}$. Bragg fringes are directed along $\langle 130 \rangle$ in $\{110\}$ sectors. (b) Images from collapsed polyethylene pyramids with Bragg fringes directed along $\langle 1\bar{3}0 \rangle$ in $\{110\}$ sectors. (c) $(hk0)$ diffraction pattern from orthorhombic paraffin crystal in space group $Pca2_1$. (d) Nonplanar crystal habit of n-C_mH_{2m+2} for $m \geqslant 44$ [17].

the orientation of diffraction contrast bands in individual sector domains (Fig. 1b). In addition, high resolution images [17] also demonstrated that the molecular packing of successive bands was similar in alkanes and polyethylene; therefore tight folds with strictly adjacent re-entry need not necessarily be invoked to

explain the observed phenomena in polyethylene. Further studies, in which n-$C_{50}H_{102}$ and n-$C_{60}H_{122}$ were co-crystallized in order to create a disordered surface, established that surface order is not a prerequisite for the phenomenon of sectorized crystallization [18]. A second important issue which was addressed in this investigation is that of chain tilt with respect to the layer normal (clearly the surface area per stem is increased by molecular tilt, thus alleviating the packing problem at the lamellar surface). This phenomenon is related to molecular symmetry, which is *mm* for an odd-numbered n-alkane and $2/m$ for an even numbered n-alkane. The former will pack preferentially in a rectangular array of untilted chains, whereas the latter is required to preserve a center of inversion in the layer packing and gives rise to oblique layers with tilted chain axes. In the case of polyethylene, it was shown that an oblique layer tilt axis occurs along the methylene subcell diagonal, corresponding to one of Kitaigorodskii's $R[\pm 1, \pm 1]$ layers [19]. The direction of this tilt axis was shown to change in each lamellar sector. It was shown that the growth of a three-dimensional lamellar habit from solution must be nucleated by a four-fold twin, where the lamellar sectors meet at the projected diagonals [18].

Poly(ether ether ketone)

Poly(ether ether ketone) (PEEK) is a high performance engineering thermoplastic, the repeat formula of which is $[C_6H_4OC_6H_4OC_6H_4CO]_n$, which has a very high melting temperature. Because of its high temperature/high strength characteristics, its exceptional chemical resistance, and its melt process ability, PEEK is extremely interesting for applications such as reinforced composites, coatings, electrical connectors, etc. In a series of articles describing electron microscopy and diffraction in PEEK, it was shown that the unit cell in the narrow fibrous crystals is orthorhombic ($a = 7.75$, $b = 5.89$, $c = 9.88$ Å) (Fig. 2) and very similar to other aromatic polymers such as poly(*p*-phenylane oxide) (PPO) and poly(*p*-phenylane sulfide) (PPS) [20]. The electron

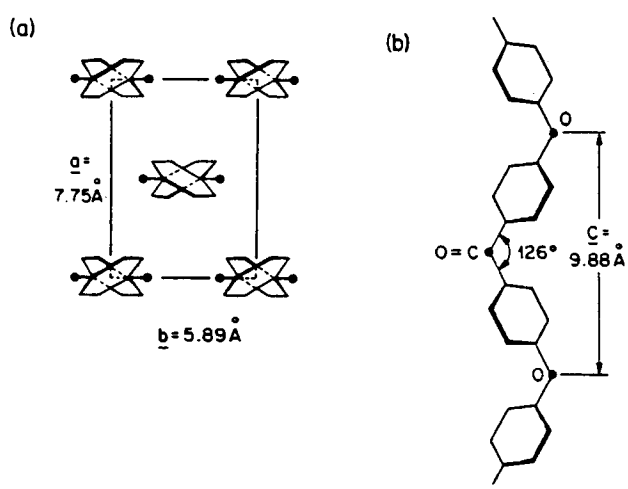


Figure 2. Projection of (a) the unit cell of PEEK along the *c* axis, and (b) the molecular conformation of PEEK along the *a* axis [20].

diffraction results were not consistent with the packing scheme obtained from X-ray results for PPO [21] and PPS [22] in which the molecules are located at the corners and center of an orthorhombic unit cell. The explanation for the puzzling double melting peak observed in these materials was subsequently shown by electron microscopy to lie in a second morphological form, having the same unit cell and *b*-axis preferred growth direction, but in this case the molecule is not parallel to the lamellar normal but inclined to it by 38.1°, leading to {102}-type fold surfaces [23].

Polypropylene

Polypropylene can be synthesized in the isotactic (iPP) and syndiotactic (sPP) forms. Electron microscopy has played a major role in establishing its structure. This includes the determination of the unit cell orientation and preferred growth direction in quadrites and spherulites of iPP [24], the elucidation of the molecular origin of lamellar branching [25], and the determination of the morphology and structure of the γ-polymorph [26]. The structure and defects of sPP were studied by using a combination of X-ray scattering, electron diffraction, and molecular modeling [27] (Fig. 3). The chain in its $(t_2g_2)_2$ conformation was shown to pack in a fully antichiral orthorhombic cell along both the *a* and the *b* axes, with $a = 14.5$ Å, $b = 11.2$ Å, and $c = 7.4$ Å. Crystallization at lower temperatures causes streaking on $\{h20\}$ spots in electron diffraction, indicating defects along the *b* axis consisting of rows of chains displaced by $b/4$. In a subsequent paper on their morphology, these crystals were shown to exhibit transverse fractures, which are associated with an order of magnitude greater thermal expansion along the crystallographic *b* axis with respect to the *a* axis. This has drastic consequences on the mechanical properties [28].

These representative examples of crystal structure and morphology obtained from three major types of macromolecules demonstrate that electron microscopy can provide crucial information about fundamental issues in polymer physics, but that these results have to be viewed in the larger framework of other methods such as X-ray and neutron scattering, differential scanning calorimetry, and density measurements, as well as Raman and infrared data. Fortunately, there has been an increasing tendency towards more collaborative research between scientists from different fields of expertise, a development which is certainly in the interest of polymer science.

13.2.1.2 Bulk Morphology

Linear polymers can crystallize with varying degrees of order. Most polymers crystallize by chain folding to form thin lamellae, the internal structure of which corresponds to that of single crystals. This central core is similar to crystals of a monodisperse oligomer. The essential difference with respect to single crystals from solution is the trajectory of the molecular chain in the noncrystalline regions, since in the bulk these serve to connect the crystalline entities with one another. As flexible molecules form Gaussian coils in the melt [1], the question arises as to how they reorganize during crystallization in order

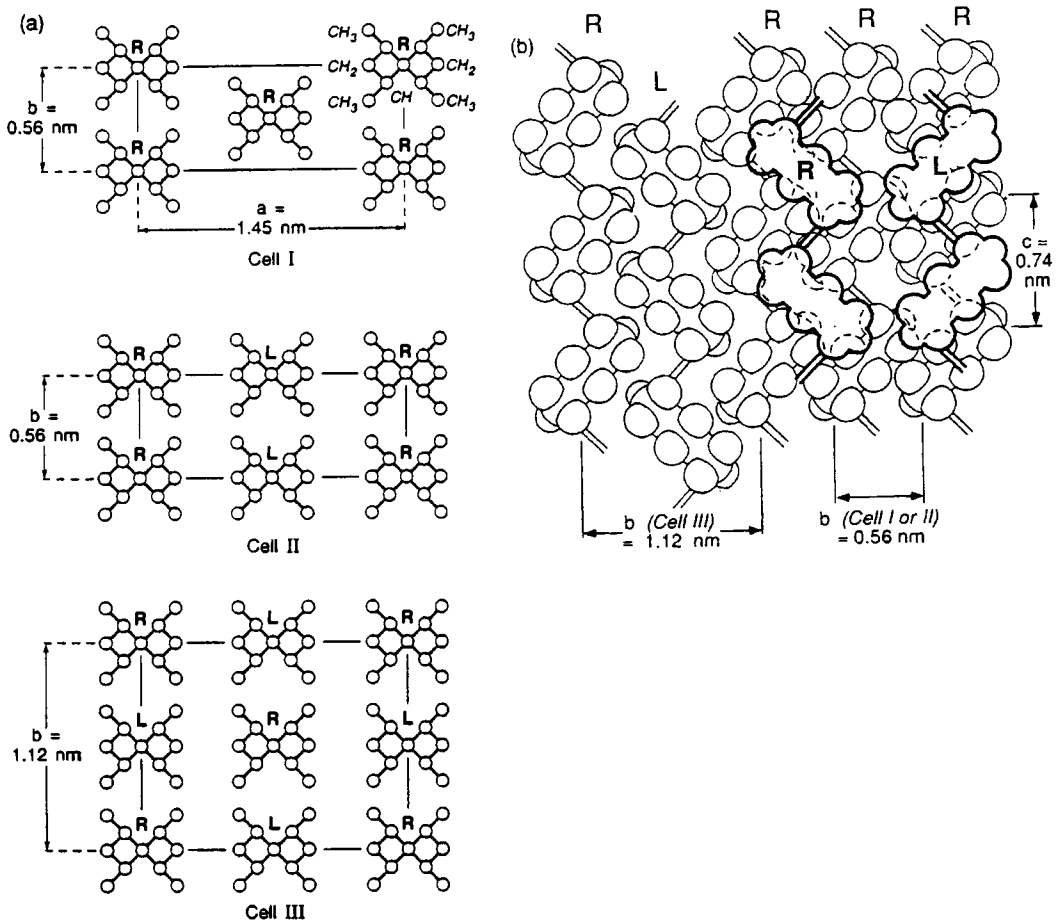


Figure 3. (a) Proposed unit cells for the stable $(t_2g_2)_2$ form of sPP, shown in the *c* axis projection. Cell I: fully isochiral cell, carbon centered. Cell II: Primitive cell with antichiral packing of chains along the *a* axis. Cell III: fully antichiral, body-centered cell. Right- and left-handed helical molecules are denoted by R and L, respectively [27]. (b) Schematic diagram showing the origins of isochiral (RRR) and antichiral (RLR) molecular packing in sPP. While the right-handed molecule can deposit only between a pair of right-handed substrate helices, the left-handed molecule can deposit only directly on top of a right-handed one. The left half shows that only antichiral deposition is possible for a cell III substrate [27].

to form a condensed phase in which lamellae with a typical thickness of about 100 Å are interspersed with amorphous regions. Depending on the crystallization conditions, the lamellae become twisted and radiate from the center to produce a spherulite: the radiating primary lamellae are interspersed with secondary lamellae which crystallize during cooling, and with amorphous chains which are located at the surfaces of both primary and secondary lamellae as well as at the crystal edges. The degree of twist and bend depends on the molecular weight and cooling rate.

Kinetic studies of the crystallization process have shown that the morphology of bulk polymers depends on the crystallization temperature and the rate of

cooling [29–31]. An analysis of crystallization rates in the vicinity of the melting temperature showed that polymer crystallization is a nucleation-controlled process. The details of the nucleation process were first analyzed by Hillig for the nucleation and subsequent growth of small molecules on previously formed substrate [32]. The conclusions can be adapted in a straightforward manner, without recourse to molecular models, for the crystallization of long-chain molecules [33–36]. Undoubtedly, the major type of superstructure observed in bulk polymers is the spherulite. For the detailed elucidation of spherulite morphology by surface replication methods, the work of Bassett represents a major contribution. The technique of permanganic etching [37] has shown the arrangement of crystallites (lamellae) in many bulk polymers, revealing screw dislocations on lamellar surfaces as well as the organization at a superstructural level. Two images of high- and low-density polyethylene prepared by using this method are shown in Fig. 4. Subsequently the superstructural organization of lamellae was studied in other technologically important polymers, notably poly(aryl ether ether ketone) [38] and isotactic poly(4-methylpentene-1). The universal feature was shown to be the crystalline lamellae with branching at, and divergence from, giant screw dislocations, leading to a space-filling (spherulitic) framework of dominant lamellae with the fastest growth direction being radial [39]. Subsequent growth occurs within this framework. The genesis of screw dislocations can be controlled by increasing the undercooling, giving rise to dislocations at the resulting interface. The asymmetry in growth terraces associated with a screw dislocation is expected to give rise to a twist because of lamellar divergence. The well-known appearance of banded spherulites was shown to be linked to the formation of isochiral screw dislocations coupled with the tendency of adjacent lamellae to diverge. The electron microscopic investigation on poly(4-methylpentene-1) combined with differential scanning

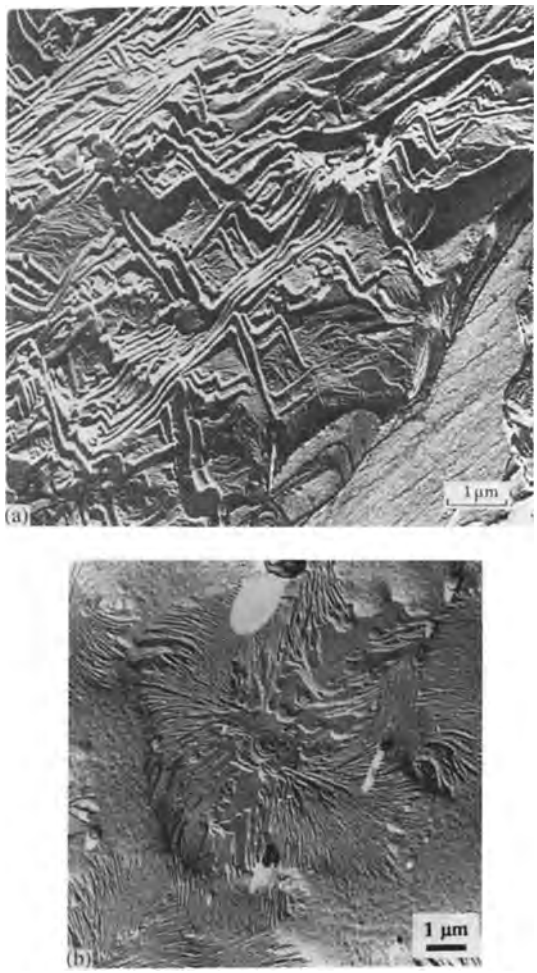


Figure 4. (a) High density polyethylene showing ridged (roof-shaped) lamellae and thin (secondary) lamellae [37]. (b) Embryonic spherulites of linear low-density polyethylene grown at 124°C [39].

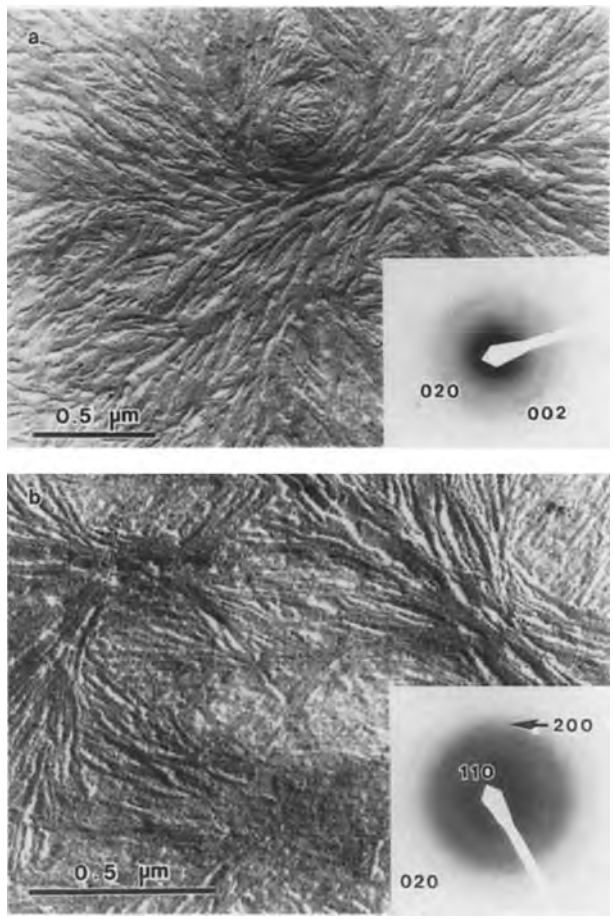


Figure 5. Bright-field transmission electron micrographs and selected-area diffraction patterns from thin PEEK films crystallized at (a) 270°C and (b) 250°C. The diffraction pattern in (a) stems from the lower left-hand region of the field, while that in (b) arises from the central region [41].

calorimetry verified earlier results on polyethylene, poly(aryl ether ether ketone)polypropylene [40], and poly(aryl ether ether ketone) (Fig. 5); that is, that the subsidiary lamellae melt at lower temperatures than the dominant lamellae [41]. Furthermore, subsidiary lamellae deform earlier during tensile drawing. In contrast to the branched copolymers, where many chain branches are expected to be located at the crystal surfaces, thus preventing isothermal crystal thickening, the linear materials have repeatedly shown this phenomenon. Electron microscopic studies on poly(4-methylpentene) revealed that, when already existent lamellae are held slightly below their melting point, local effects at lamellar edges lead to continuous lamellar thickening, thus achieving higher stability and melting points.

While surface replication methods such as these give detailed qualitative information about the morphology and superstructural arrangement of polymer crystallites within the spherulites, a quantitative assessment of crystal growth became increasingly essential as fundamentally different concepts regarding the nucleation and growth of polymer crystallites, as well as their surface morphologies, gave rise to apparently

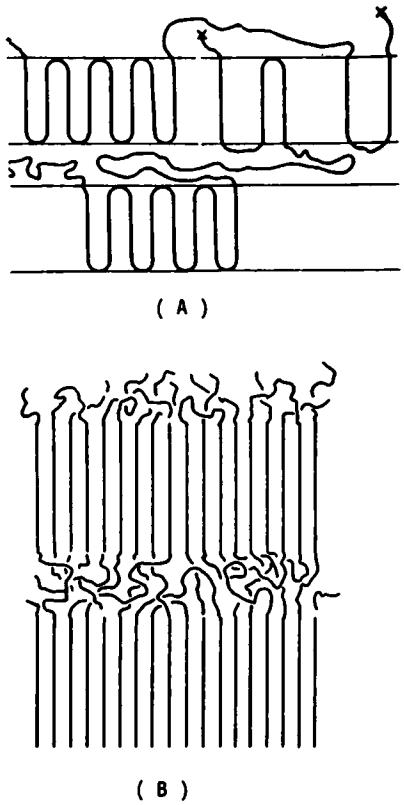


Figure 6. Two examples of proposed models for the structure of the amorphous regions in semicrystalline polymers: (a) central core model with eight adjacent re-entry folds in the cores; (b) 'switchboard' model [42].

insoluble controversies. The sketch shown in Fig. 6 summarizes the molecular trajectory problem [42] which is concerned with the question of whether or not the molecule emerging from the crystal surface returns to the same crystal in an adjacent position with tight folding. The question relating to the number of molecules between crystallites and their conformation in the non-crystalline sequences has direct bearing on: the strength of polymers; their fracture, crazing, and yield behaviors; the glass transition temperature; diffusion; and physical aging [43–45].

For polymeric molecules the temperature interval just below the equilibrium melting temperature is a metastable zone in which nuclei do not form at a detectable rate but in which crystals, once formed, can grow (regime I). Below this metastable temperature zone, nuclei may form spontaneously, either homogeneously or heterogeneously (regime II). As the substance cools further, a high-viscosity zone is reached where the formation of nuclei is again inhibited and growth is again very slow (regime III). Studies of the overall crystallization of spherulites in linear polyethylene over an extended molecular weight range showed that the temperature coefficient of the overall crystallization rate increases by a factor of 2 during the transition from regime I to regime II [46, 47].

From these considerations, and from experimental light scattering evidence, it became clear that the supermolecular organization of lamellae depends on both the undercooling and the molecular weight in a complex manner. This was first demonstrated for linear polyethylene in a light-scattering study by Maxwell and Mandelkern [36] (Fig. 7) and subsequently for branched polyethylene using molecular fractions [48]. A detailed analysis of the crystallite morphologies underlying these superstructures in different temperature and molecular weight regions was undertaken in a series of electron microscopy studies of thin microtomed sections stained with chlorosulfonic acid and uranyl acetate, a method first proposed by Kaning [49]. Typical morphologies are shown in Fig. 8 for linear polyethylene. Roof-shaped and curved lamellae are observed for low and high undercoolings, respectively [50, 51]. It is clear that these

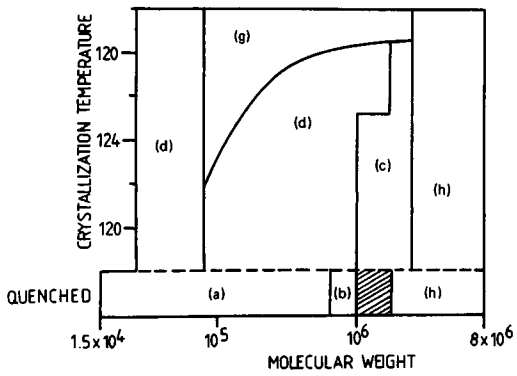


Figure 7. Schematic of the observed morphology in linear polyethylene as a function of molecular weight and crystallization conditions. The letters in parentheses refer to the morphologies: (a) spherulites consisting of curved lamellae; (b) spherulites consisting of twisted lamellae; (c) poorly developed spherulites; (d) large lamellar sheets; (g) lamellar sheets with multiple corrugations; (h) thick curved lamellae interspersed by thin lamellae [36].

features are identical to the ones described by Bassett on the basis of surface replica studies performed at about the same time. However, staining also revealed three important parameters, namely the thickness of the amorphous and crystalline regions as well as the molecular tilt with respect to the crystal surfaces. The images immediately suggest that the lamellar morphology and thickness of the (black) amorphous regions is related to the molecular trajectory, and this in turn is related to the molecular weight and undercooling. When the crystal tilt angle (not to be confused with the above molecular tilt) with respect to the electron beam is optimized for maximum contrast, the data can be quantified and the dependence of the crystal thickness d_c plotted as a function of molecular weight for the linear material (Fig. 9). These investigations thus revealed details about lamellar thickness as a function of molecular weight and crystallization temperature which were crucial for the understanding of crystallization kinetics and mechanical properties [52].

The electron micrographs also provided the information about the molecular tilt with respect to the lamellar normal in linear bulk material (as determined from the apex angle of the roof-shaped lamellae) and showed that the fold surfaces are predominantly $(\bar{1}01)$ in regime I, but $(\bar{2}01)$ and $(\bar{3}01)$ in regime II [53]. The results clearly demonstrated that the tilt angles increase with larger undercoolings because this alleviates the problem of dissipating the flux of chains emanating from the basal plane of the crystallites and that this phenomenon is related to the transition from regime I to regime II crystallization kinetics.

For branched co-polymers containing short branches at typical mole fractions of a few %, numerical analysis of electron micrographs and Fourier transformed small-angle X-ray scattering results provided the information required to formulate a new model of melting and crystallization for ethylene copolymers, which was shown to be quite different from the mechanism taking place in linear systems [54, 55]. In contrast to linear polyethylene, where lamellae continue to thicken at high temperatures, branched polyethylene shows a reversible change in long spacing related to the amorphous regions, while the crystallites retain their thickness, which is in fact related to the average distance between branches so that branches are located preferentially at the crystal surfaces [31, 52], thus preventing crystal thickening (Fig. 10).

The possibility that numerical values in stained thin sections may be falsified by beam damage or microtoming should

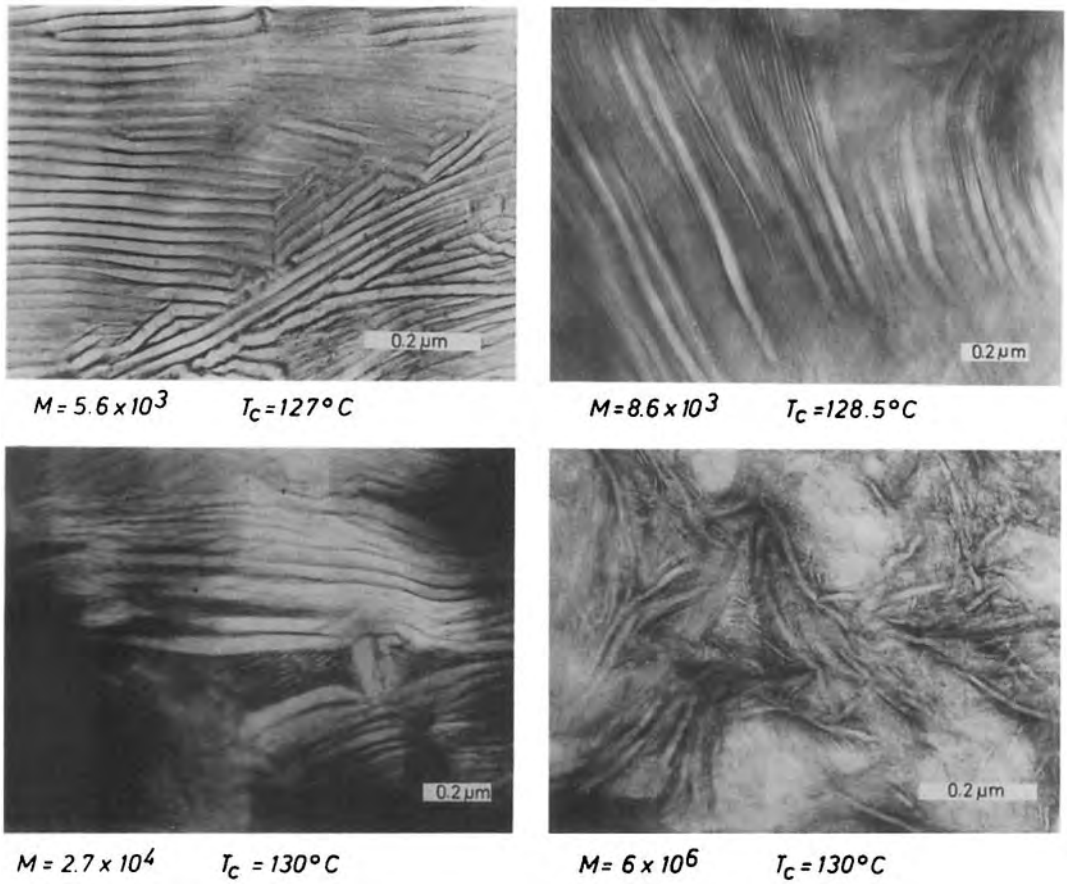


Figure 8. Lamellar morphology of linear high density polyethylene as a function of crystallization temperature and molecular weight [50, 52].

always be recognized and can be avoided by (1) proper attention to experimental procedures and (2) quantitative comparison with the results of X-ray small angle scattering [54] and Raman longitudinal acoustic mode measurements [56, 57] (Fig. 11).

In view of its enormous strength in providing detailed numerical information, it would be very short-sighted to reject quantitative analysis of electron micrographs because of difficulties (which can be overcome). Undoubtedly the insight into lamellar morphology provided by numerical electron microscopic investigations has also contributed to the understanding of polymer structure which has made the development of new materials with improved physical properties possible.

13.2.1.3 Defects in Bulk Polymers

Long before high resolution methods became generally useful in the investigation of polymers, electron microscopy played a crucial role in the characterization of polymer defects, which are of course related to the specific nature of the intermolecular and intramolecular

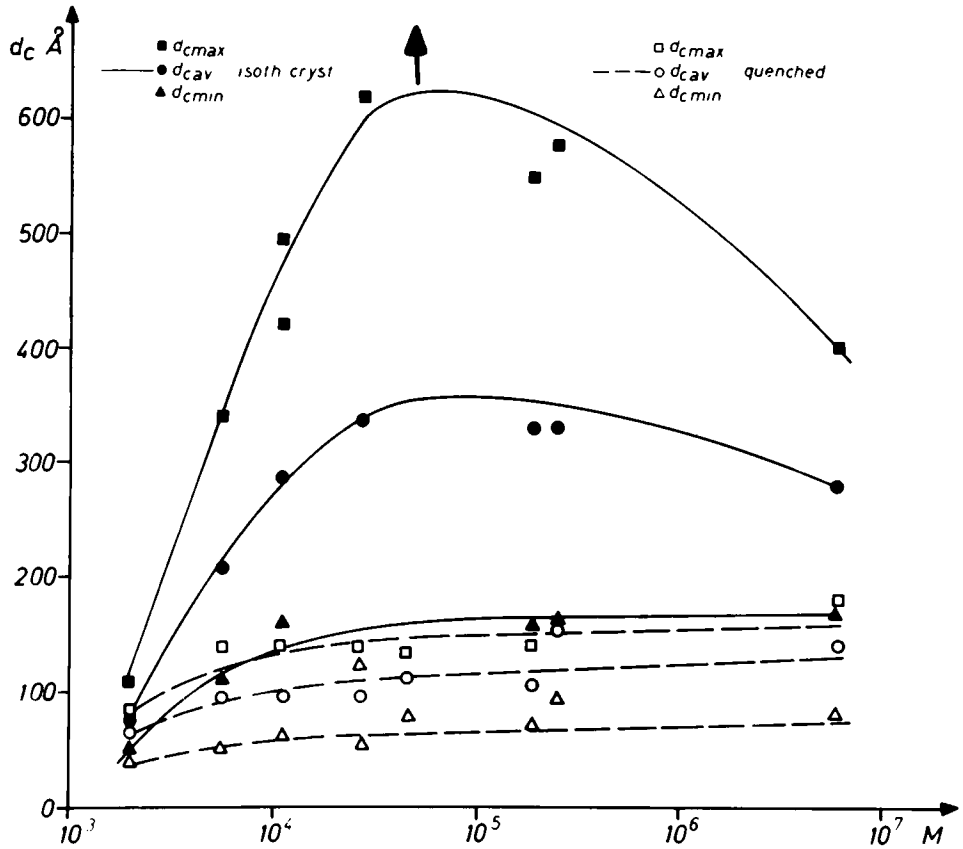


Figure 9. Dependence of crystal thickness on molecular weight [52].

bonds, local defects in bonding, chain branches, chain ends, etc. The analysis of such defects has been given impetus by the necessity to understand their effect on the mechanical, electrical, and optical properties as well as on the phase transitions and surface properties of polymers. In addition, the deliberate manipulation of defects offers the possibility of creating and controlling specific microstructures in ordered polymers [58]. Specific defects which have been identified by electron microscopy include elongated point defects, line defects, and grain boundaries.

Elongated point defects

Local translations and rotations of the polyethylene chain give rise to a family of single chain point defects [59].

Line defects

Line defects in polymers include dislocations (singularities in translational order), disclinations (singularities in orientational order), and dispirations (singularities in both translational and orientational order).

Screw dislocations, which have been associated with lamellar single crystal growth, have been identified in a careful

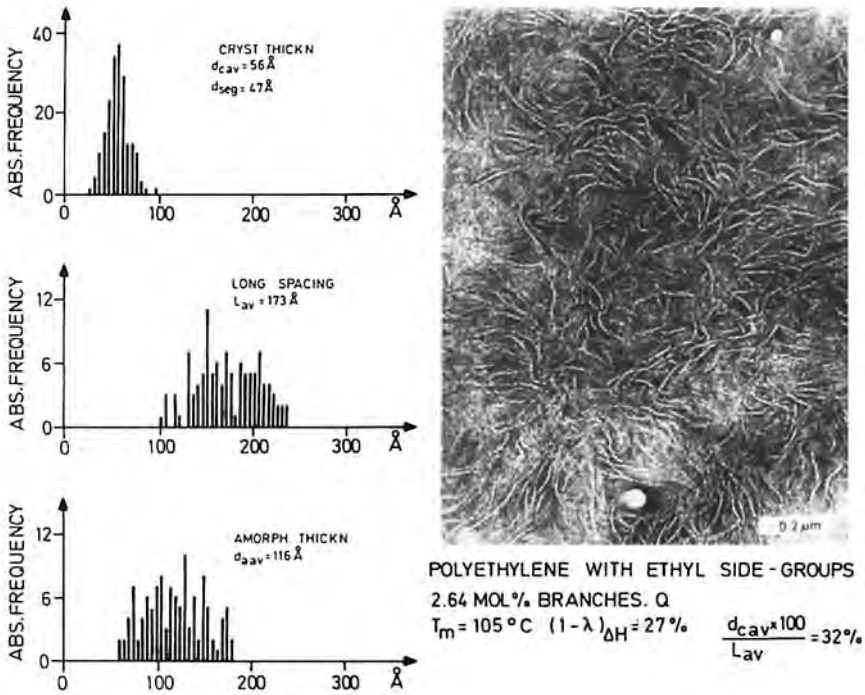


Figure 10. Polyethylene with ethyl side-groups: 2.64 mol% branches, $QT_m = 105°C$, $(1 - \lambda)_{\Delta H} = 27\%$; $(d_{cav} \times 100)/L_{av} = 32\%$ [52].

study as a four-fold twin in which lamellar sectors meet at the projected diagonals [60].

Edge dislocations in diacetylenes have been identified by etching experiments [61].

Aggregations of dislocation have been shown to create grain boundaries which can accommodate tilt or twist between domains [62].

The molecular organization of liquid crystalline polymers has been studied for $s = \frac{1}{2}$ disclinations [63] and for $s = 1$ disclinations [64] where s is the disclination strength.

Grain boundaries

Grain boundaries in extended chain crystalline polymers have been discussed extensively by Martin and Thomas [65]. In addition to these, specific defects related to copolymers are discussed in the following section.

13.2.2 Copolymers

When linear chains consist of different kinds of repeating units, A and B, they are called 'copolymers'; in this case either one or both of the components are amorphous. There are four extreme possible arrangements:

A–B–A–B–···	Alternating copolymer
A–A–A–A–B–B–B–A–A–A–A–	Ordered copolymer
···A–B–A–A–B–B–B–A–B–B–	Random copolymer
···–A–A–A–A–A–A–A–	Graft copolymers
	B B
	B B
	B B
	B B

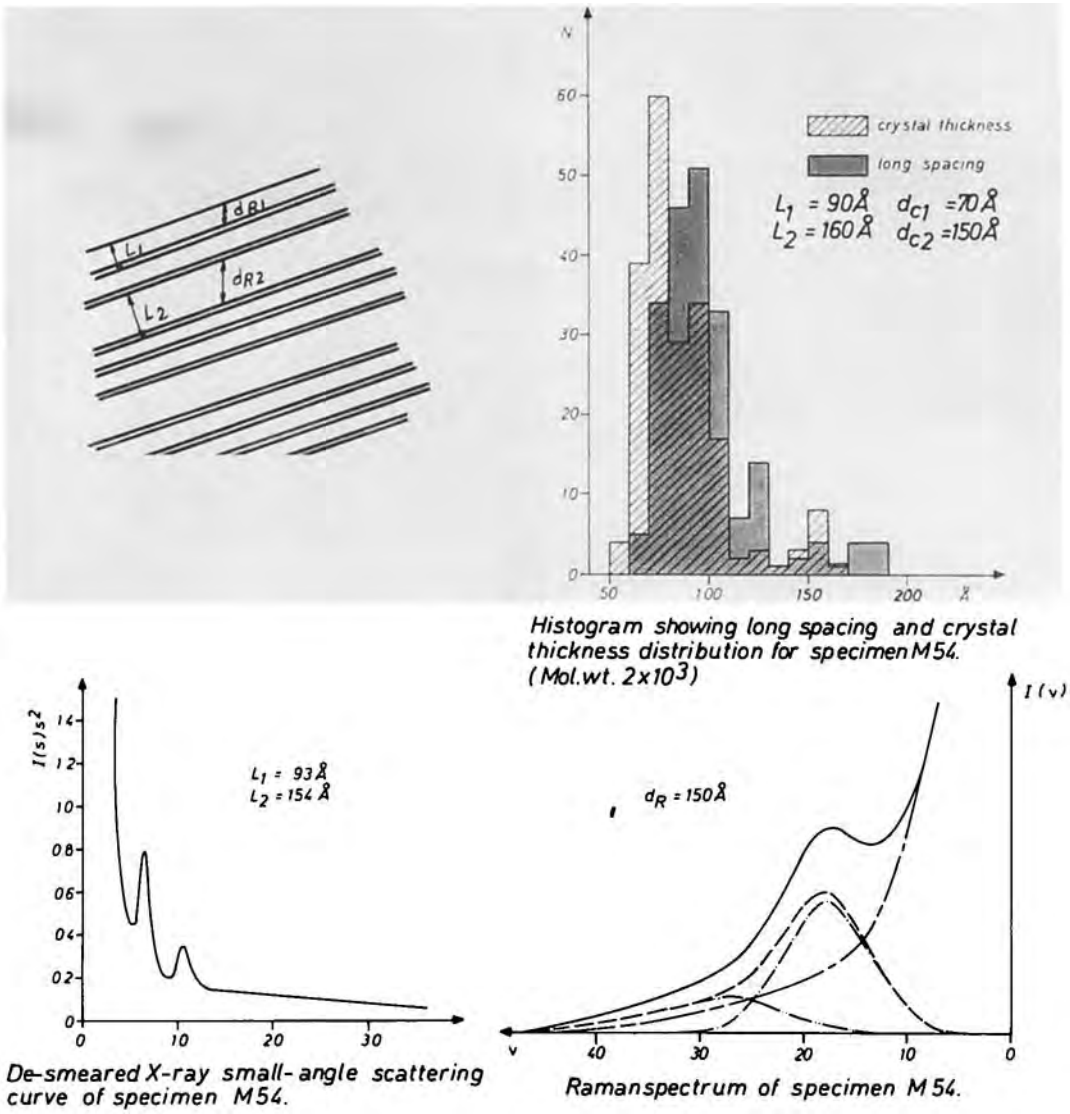


Figure 11. Comparison between Raman, Electronmicroscopy and small angle x-ray scattering for binodal distribution of lamella thickness and long spacing distribution in polyethylene. Molecular weight = 2×10^3 Da; $T_c = 117.3°C$ [57].

The different arrangements and types of copolymer unit give rise to very rich phase diagrams with an enormous range of superstructures, each of which gives rise to different physical properties. Here the emphasis usually lies on the optimization of mechanical properties, but it is quite feasible that specific optical or dielectric properties will be required. The chemistry and physical behavior of block copolymers have been reviewed in several comprehensive books [66, 67].

13.2.2.1 Equilibrium Morphologies

Theoretical considerations for the case of symmetric AB diblock copolymers consisting of two kinds of polymer A and B with the same degree of polymerization have been pioneered by Helfand [68] for the strong segregation limit and by Leibler for the weak segregation limit (WSL) [69]. For the WSL, Leibler has used the mean-field theory to explain the emergence of lamellar structures. Monte Carlo simulations on the ordered structures of bulk systems have been reported by Fried and Binder [70].

The complexity of the superstructures for more complicated systems makes electron microscopy a mandatory tool for their investigation and a basis for model calculations. Some possible staining agents are OsO_4 and RuO_4, depending on the components involved. Melts of AB diblock copolymers form phase-separated structures in which domains of one phase are separated from those of the other by an interfacial region. Early theories predicted three stable, ordered morphologies in both weak and strong segregation limits: alternating lamellae of A and B, hexagonally packed cylinders of A in matrix B, and spheres of A in matrix B [71].

Electron microscopic studies of star block copolymers identified a new 'equilibrium' microdomain morphology in styrene–isoprene copolymers, namely the ordered bicontinuous double-diamond (OBDD) [72]. The basic unit of the structure is a tetrahedral arrangement of short polystyrene rods (Fig. 12). Such units are interconnected on a cubic lattice with *Pn3m* space group symmetry. The resultant structure consists of two translationally displaced, mutually interwoven but

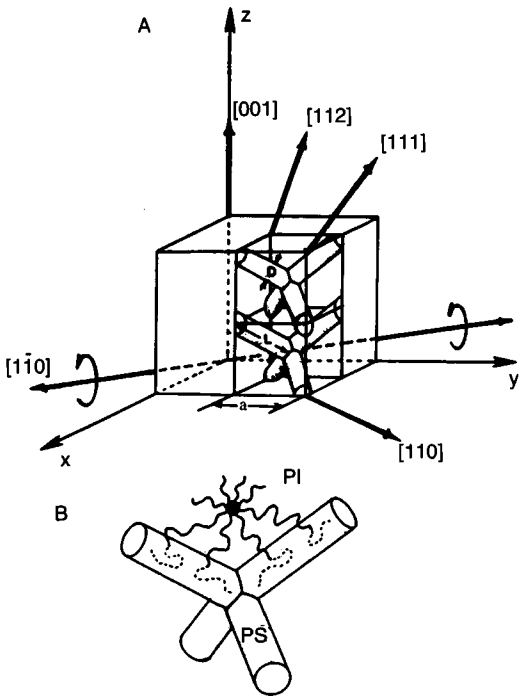
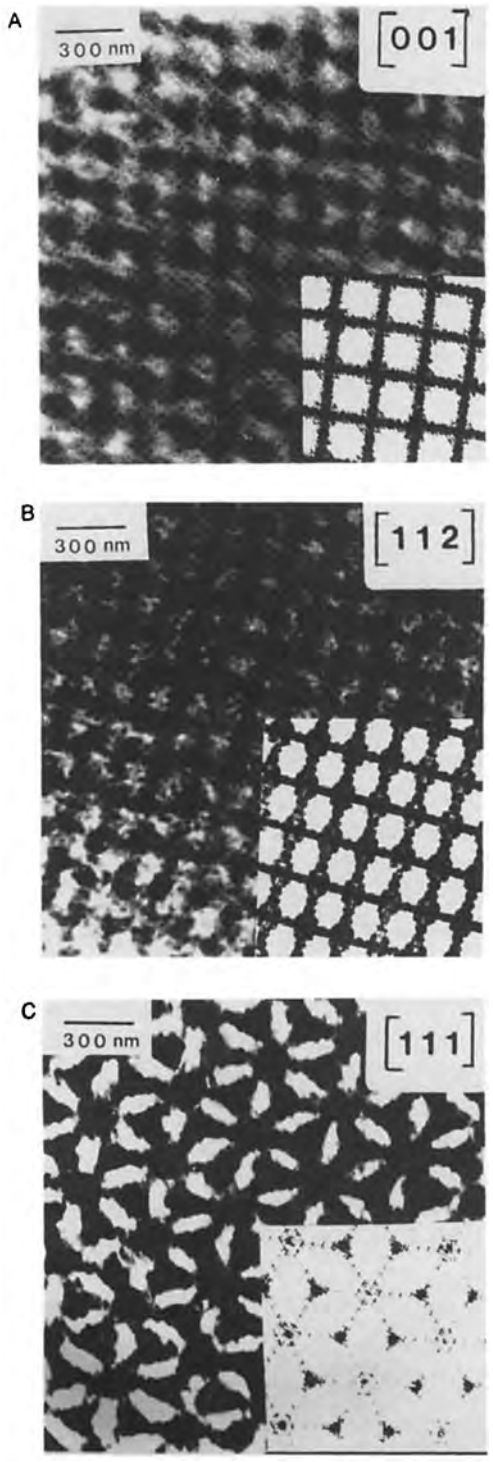


Figure 12. (a) Schematic of the OBDD structure showing the arrangement of the basic tetrahedral units. These interconnect to form two interwoven, but independent, diamond networks of polystyrene rods. (b) Schematic of the arrangement of the polystyrene and polyisoprene chains in the OBDD domain space [72].

unconnected three-dimensional networks of polystyrene rods embedded in a polyisoprene matrix. By comparing computer-generated images with experimental images, it was shown that each of the separate polystyrene networks has the symmetry of a diamond cubic lattice (Fig. 13).

Recently, methods have been developed for calculating the free energy of bicontinuous morphologies in the strong segregation limit [73]. In the weak segregation regimes, a variety of new structures have been identified; the lamellar-catenoid [74, 75], hexagonally modulated lamellae, hexagonally packed lamellae [76], and the



double gyroid [77]. The analysis of such morphologies makes tilting series and a comparison with X-ray scattering mandatory, as the following example from Hadjuk et al. [77] indicates. Thin sections of a polystyrene–polyisoprene copolymer were microtomed at −100°C, stained with OsO_4, and examined in different projections in a 200 kV JEOL 200 CX electron microscope. Two prominent projections indicated a cubic and a hexagonal geometry. Selected regions of the images were then digitized, a two-dimensional Fourier transform performed, and the geometry in the images deduced as [100] (Fig. 14) and [111] projections of a cubic structure. Parallel to this, small angle X-ray experiments on the same sample revealed reflections at position ratios $\sqrt{3}:\sqrt{4}:\sqrt{7}:\sqrt{8}:\sqrt{10}:\sqrt{11}:\sqrt{12}:\sqrt{13}:\sqrt{15}:\sqrt{16}:\sqrt{19}$, consistent with an *Ia3d* space group. Subsequently, computer models of structures with space groups consistent with the observed reflection ratios and extinctions were simulated in both direct and reciprocal space from different projections and used to analyze the TEM images and diffraction patterns. The theoretical concepts behind the model structures are based on the creation of minimal surfaces. These are two-dimensional constructs embedded in three dimensions which possess zero mean curvature (ZMC) at all points on the surface, as proposed for bicontinuous fluids [78], or constant mean curvature (CMC), which minimize the area among all nearby surfaces produced by perturbations which

Figure 13. Images of styrene–isoprene copolymers showing the (a) [001], (b) [112], and (c) [111] projections, and the corresponding computer-generated images [72].

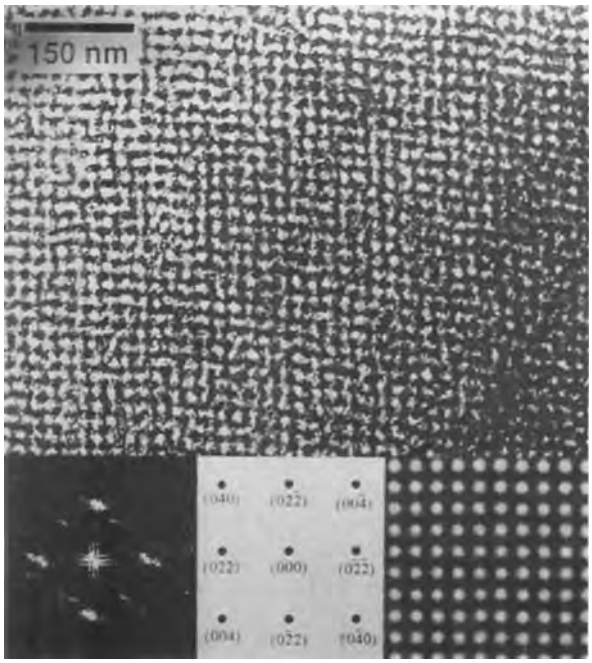


Figure 14. TEM micrograph of polystyrene–polyisoprene copolymer showing approximately four-fold symmetry corresponding to the [100] projection of the cubic structure. The insets show the image fast Fourier transform and indexed diffraction pattern. The inset on the right-hand side shows a simulated [100] projection of the constant-thickness gyroid structure [77].

preserve the volumes of regions on either side of the surface [79, 80]. The ordered bicontinuous double-diamond (space group *Pn3m*) and double gyroid (space group *Ia3d*) are examples of such structures. Finally, the model was viewed from the projections corresponding to the electron micrographs, and the Fourier transform calculated for comparison with the small angle X-ray scattering results. The results of the calculations for the gyroid structures are shown in Fig. 15.

In symmetric poly(A-block-B-block-C)triblock copolymers, electron microscopy has revealed a very rich phase diagram consisting of a multitude of new morphologies such as spheres and rings located at the surfaces of cylinders, multiple and single helical strands around cylinders (Fig. 16). In addition undulating sheets and cylinders (Fig. 17) embedded in a matrix of the major component [81–83] were characterized. An understanding of morphological features observed in these triblock copolymers is based on the following simplified considerations.

The free energy per chain (F_c/kT) in the strong segregation limit can be written as:

$$\frac{F_c}{kT} = \frac{F_c^{surf}}{kT} + \frac{F_c^{elastic}}{kT}$$

where F_c^{surf} is the surface free energy and $F_c^{elastic}$ is the elastic contribution

$$\frac{F_c^{surf}}{kT} = \gamma_{AB} \cdot \sum_{AB} + \gamma_{BC} \cdot \sum_{BC} + \gamma_{AC} \cdot \sum_{AC}$$

where $\sum_{xy}$ are the common contact areas of components x and y per chain experiencing a surface tension γ_{xy}

$$\frac{F_c^{elast}}{kT} = K_A\left(\frac{L_A^2}{N_A a_A^2}\right) + K_B\left(\frac{L_B^2}{N_B a_B^2}\right) + K_c\left(\frac{L_c^2}{N_c a_c^2}\right)$$

where N_A, N_B, and N_C are the number of

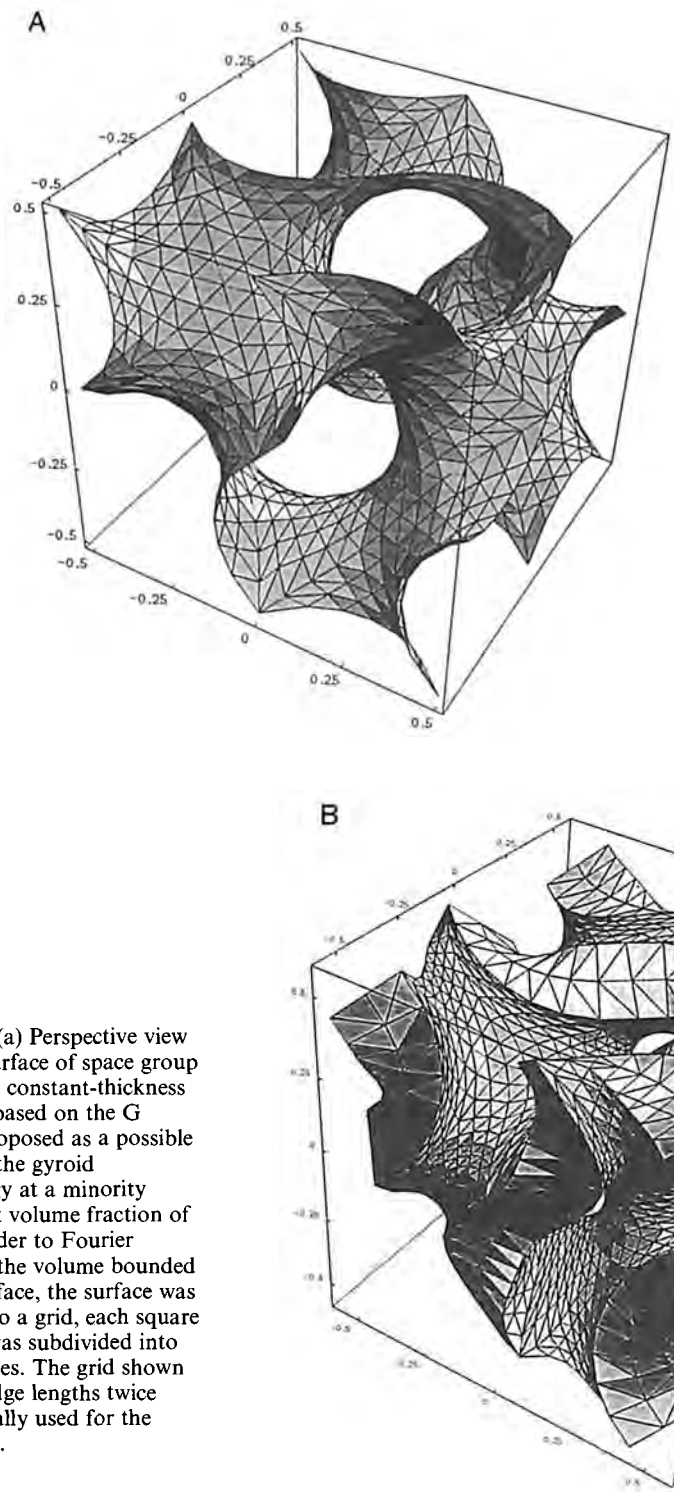


Figure 15. (a) Perspective view of the G surface of space group Ia^3d. (b) A constant-thickness structure, based on the G surface, proposed as a possible model for the gyroid morphology at a minority component volume fraction of 0.33. In order to Fourier transform the volume bounded by this surface, the surface was divided into a grid, each square of which was subdivided into two triangles. The grid shown here has edge lengths twice those actually used for the calculation.

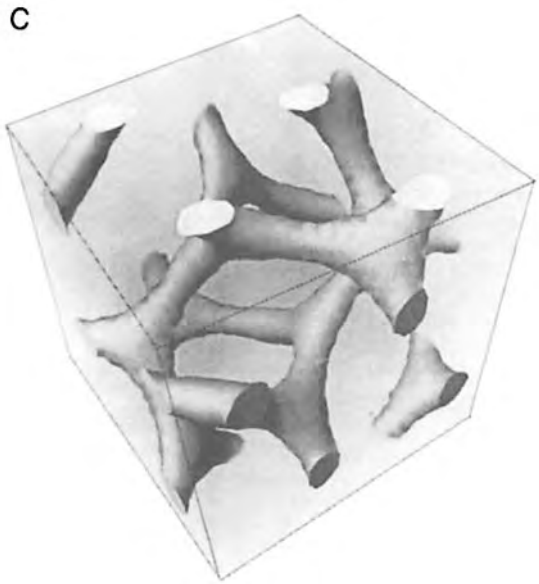


Figure 15. (c) The channel structure (minority component domains) of the constant-thickness model for the gyroid morphology [77].

monomer units of length a_A, a_B, and a_C, and L_A, L_B, L_C characterize a stretching of subchains A, B, and C, and K_A, K_B, and K_C are the force constants of the different subchains. L_A, L_B, L_C and the K values depend on the morphology of the system. The stability limit for different morphologies is obtained by minimization of the free energies. The calculations show that the morphologies which are observed by electron microscopy can be adequately predicted by theoretical arguments, and that morphological transitions can occur in ABC triblock copolymers by changing the thermodynamic balance even in the strong segregation limit. This offers an opportunity to tailor new mesoscopic microphase morphologies with specific physical properties by varying the degree of incompatibility between the components. Such ABC triblock copolymers are not only important because of the new microphase morphologies together with their associated unique mechanical properties, but also because of their potential as compatibilizers in polymer alloys [84].

13.2.2.2 Nonequilibrium Morphologies in Diblock Copolymers

Microphase separated equilibrium morphologies, such as spheres, cylinders, and lamellae are arranged within grain boundaries with very specific local arrangements (Fig. 18) [85]. The specific structure of the nonequilibrium boundaries has considerable impact on the physical properties such as elastic modulus and gas permeability, because these depend on the continuity of individual phases across the boundary. While considerably experimental [86] and theoretical [87–91] effort has been invested in the study of the equilibrium order–disorder transition and morphology versus composition phase diagrams, the grain boundary structure has only received attention more recently. In a series of articles grain boundary morphologies have been characterised in compositionally symmetric diblock copolymers forming lamellar morphologies (Fig. 19) [85].

Grain boundaries in three-dimensional crystalline materials such as metals and ceramics have five degrees of twist–tilt freedom [92, 93] of which only two, namely tilt and twist, are important for describing the grain boundaries in lamellar block copolymers, a system with only one-dimensional symmetry. The modeling approach is based on surfaces of constant mean curvature. The geometry of these interfaces was investigated by comparing experimental TEM images with ray tracing

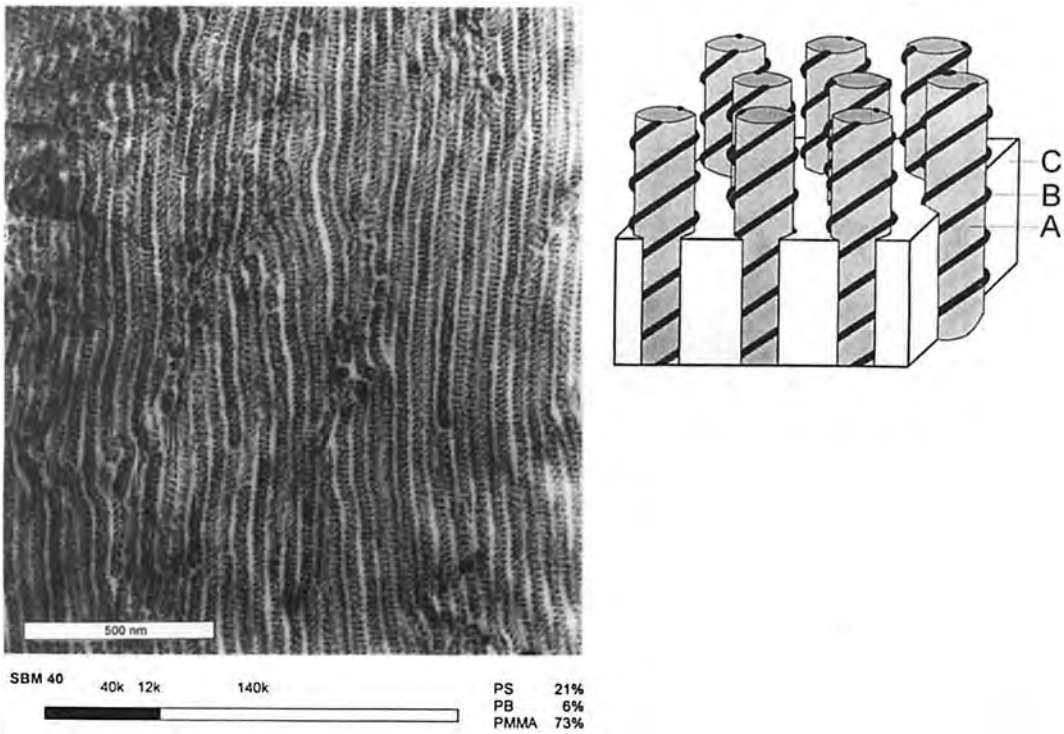


Figure 16. TEM micrograph of a triblock copolymer showing the twisted helices of polybutadiene (B) around polystyrene rods (A) in a polymethylmethacrylate matrix (C) for the compositions indicated [81].

computer simulations of the model surfaces as the projection direction was systematically varied. The two types of twist grain boundary which were observed by TEM were shown to have intermaterial dividing surfaces that approximate either Scherk's first surface or a section of a right helicoid (Figs 20 and 21) [94]. The helicoid section boundary was observed at low twist angles (<15°), while Scherk surface boundary morphologies, consisting of a doubly periodic array of saddle surfaces, was found over the entire twist range from 0° to 90° [85]. As the twist angle approaches zero degrees, the Scherk surface grain morphology transforms into a single screw dislocation which has an intermaterial dividing surface with the geometry of a single helicoid. TEM images showed the detailed core structure of this screw dislocation, and demonstrated that in lamellar diblock copolymers the screw dislocation core is nonsingular.

13.2.3 Blends

Polymer blends comprise a large class of materials in which different macromolecules are physically mixed, rather than chemically linked as in block copolymers. Generally, polymers do not mix homogeneously in solid solutions, but the phenomenon has been shown to occur in over 50 mixtures and has been investigated

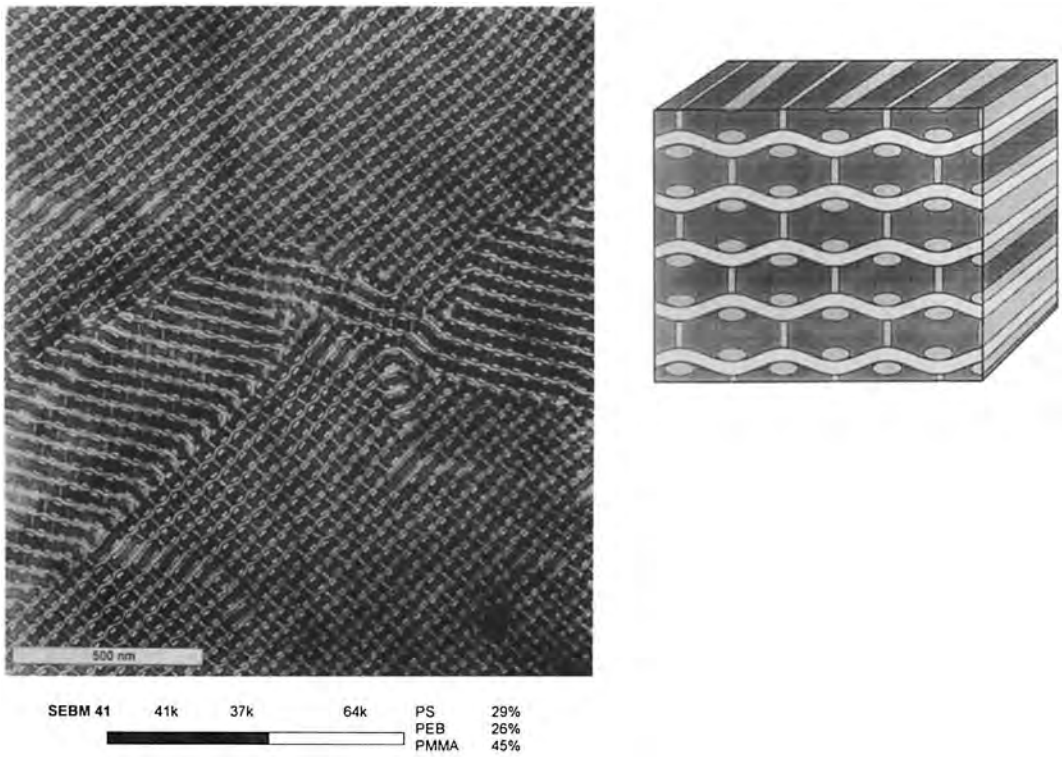


Figure 17. TEM micrograph of a triblock copolymer polystyrene–polyethylbutene–polymethylmethacrylate for the compositions indicated [82].

extensively both experimentally and theoretically [96]. Miscibility is generally limited to a certain temperature and one composition range; outside this range, the polymers phase segregate either by spinodal decomposition or by nucleation and growth.

When polymers form blends, they generally exhibit a lower critical solution temperature, as indicated in Fig. 22. The solid curves represent the cloud point (binodal) at which temperature and concentration the partial derivative of the Gibbs free energy G with respect to concentration is 0, that is $\partial G/\partial \Phi = 0$. The dotted curve encloses the spinodal region, where the transition from binodal to spinodal occurs and $\partial^2 G/\partial \Phi^2 = 0$. From the point of view of electron microscopy, the three regions are characterized by very different morphologies: (1) homogeneous outside the solid line; (2) spherical nuclei of varying sizes within a homogeneous matrix between the solid and dotted curves; and (3) a continuous, interconnected spinodal region of uniform size in a continuous matrix within the dotted region. Therefore, if the bulk samples can be quenched quickly enough from the appropriate regions, the phase diagram can be established by electron microscopy. This can be achieved in the case of long-chain polymers because the reorganization processes are sufficiently slow; however, the initial stages, which are very interesting theoretically, are generally inaccessible.

Figure 18. TEM micrograph showing the grain structure in a lamella forming poly(styrene-b-butadiene) diblock copolymer: PS 81 000 g mol^{-1}; PB 74 400 g mol^{-1} (designated SB 80/80). The inset shows a schematic illustration of the three-dimensional grain structure [85].

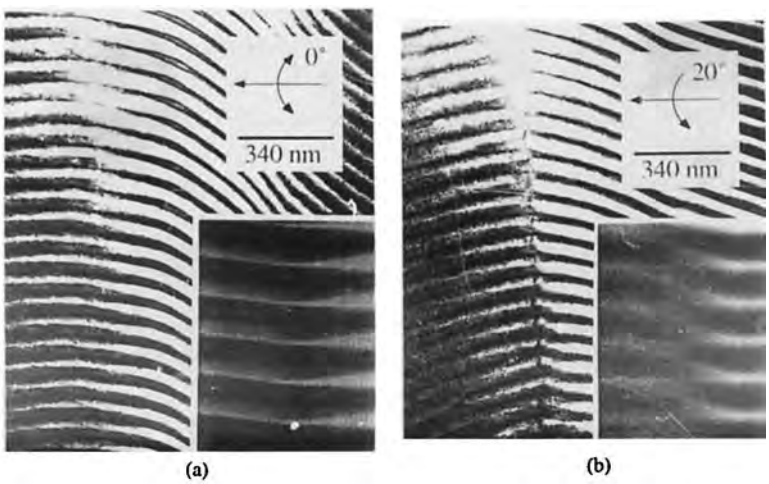


Figure 19. TEM tilt series of a helicoid section twist grain boundary (~10°) in SB 80/80. (a) The boundary and corresponding ray tracing computer simulation (inset) are viewed from a 0° tilt projection direction. (b) The specimen in the microscope and the corresponding computer simulation have been tilted by 20° about the axis shown. This tilting produces similar changes in the images of the specimen and the simulation [85].

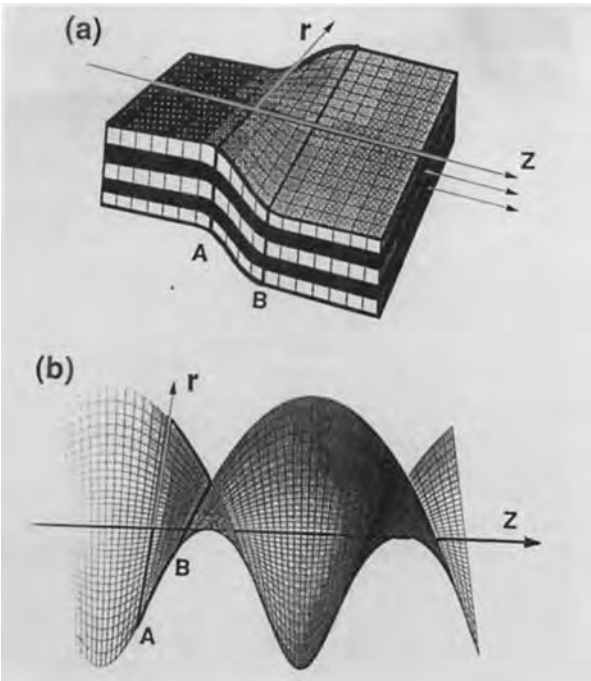


Figure 20. Helicoid section twist grain boundary morphology. (a) Lamellar material with a low angle twist grain boundary. The lamella undergo a slight 'ripple' in this boundary region which is concentrated between positions A and B. (b) The IMDS in the grain boundary region is modeled with the boldly outlined section of the right helicoid minimal surface, which is also delimited by A and B. The z axis of the helicoid, as on the mathematical surface (b), is repeated for each helicoid section of the stack of IMDSs in (a). (Computer graphic of right helicoid by J. T. Hoffman of GANG [85].)

The commercial advantage of polymer blends is due to their enormous versatility. With polymer blends in the two-phase region, a desired compromise between the properties of the two constituents may often be achieved. Typical examples include the improvement of stress-crack resistance, processability, friction, and wear resistance, as well as flame retardance. Alternatively, if one component is soluble, materials with specific pore-size distributions can be produced.

Theoretical aspects relating to the phase separation and critical behavior of polymer blends have been treated extensively in the literature [97–100]. From a theoretical point of view, the statistical mechanics of these complex fluids pose very exciting problems. The Ginsburg criterion affirms that the behavior of polymer blends becomes mean-field like in the long-chain limit [101–103]. Recently, there have been several new theoretical approaches which incorporate microscopic model features to account for chemical details: integral equation theories [104], cluster variational models [105], the lattice cluster approach [106], and the PRISM theory [107, 108].

Comparing theory with experiment is very difficult because an analytical treatment of the chemical microstructure is not possible and because the experiments can be performed over only a limited compositional and temperature range. The experimental methods of choice have generally been X-ray small angle scattering, neutron scattering, and light scattering [109, 110]. These methods have the advantage that an analysis of the scattered intensity enables important parameters such as the critical temperature to be determined, and the compositional fluctuations leading to the

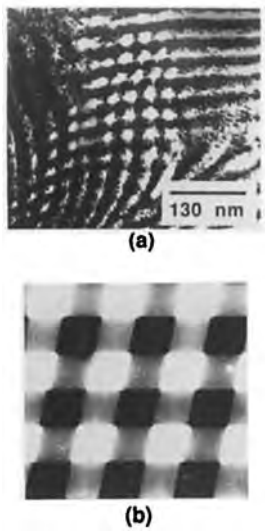


Figure 21. 80° Scherk twist boundary in SB 20/20. (a) TEM micrograph with a *z*-axis projection direction. (b) Corresponding ray tracing simulation. Three levels of intensity are present in both the TEM image and the simulation. Black regions result from the matching of PB lamella across the boundary, white regions result from the matching of PS lamella across the boundary, and gray regions result when PS and PB lamella come into contact and are separated by a Scherk saddle IMDS [85].

critical point can be detected. However, electron microscopy can play a useful role in determining the phase diagram. A typical spinodal region in tetramethyl polycarbonate/polystyrene obtained by permanganic etching of a bulk polymer surface [111] is shown in Fig. 23. In a more detailed analysis of polystyrene/polyvinylmethyl ether, the kinetics of phase segregation were studied [112, 113]. In the latter investigation, contrast was produced by mass loss of one component during exposure to the electron beam. The size d of phase-separated domains in the spinodal regions was shown to grow with time according to the relationship [113] (Fig. 24):

$$d \propto t^{\phi}$$

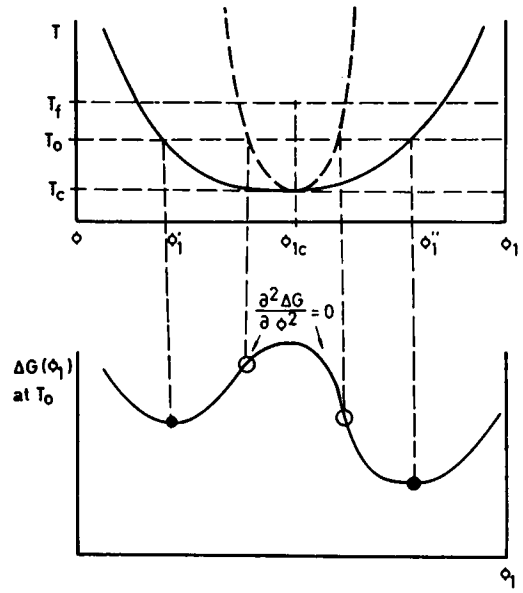


Figure 22. Phase diagram of a binary system with a lower critical solution temperature [113].

and ϕ was determined to be 1 in an intermediate region, corresponding to the theoretical expectation in that regime [114–119].

13.2.4 Fibers

The production of high strength fibers plays an essential role in polymer technology and their morphology has been studied extensively by electron microscopy. Many processing routes are possible: two important ones are gel spinning and self-assembly.

13.2.4.1 Conventional Polymers

Conventional flexible chain polymers are extended to highly extended fibers by gel spinning [120], hydrostatic extrusion [121],

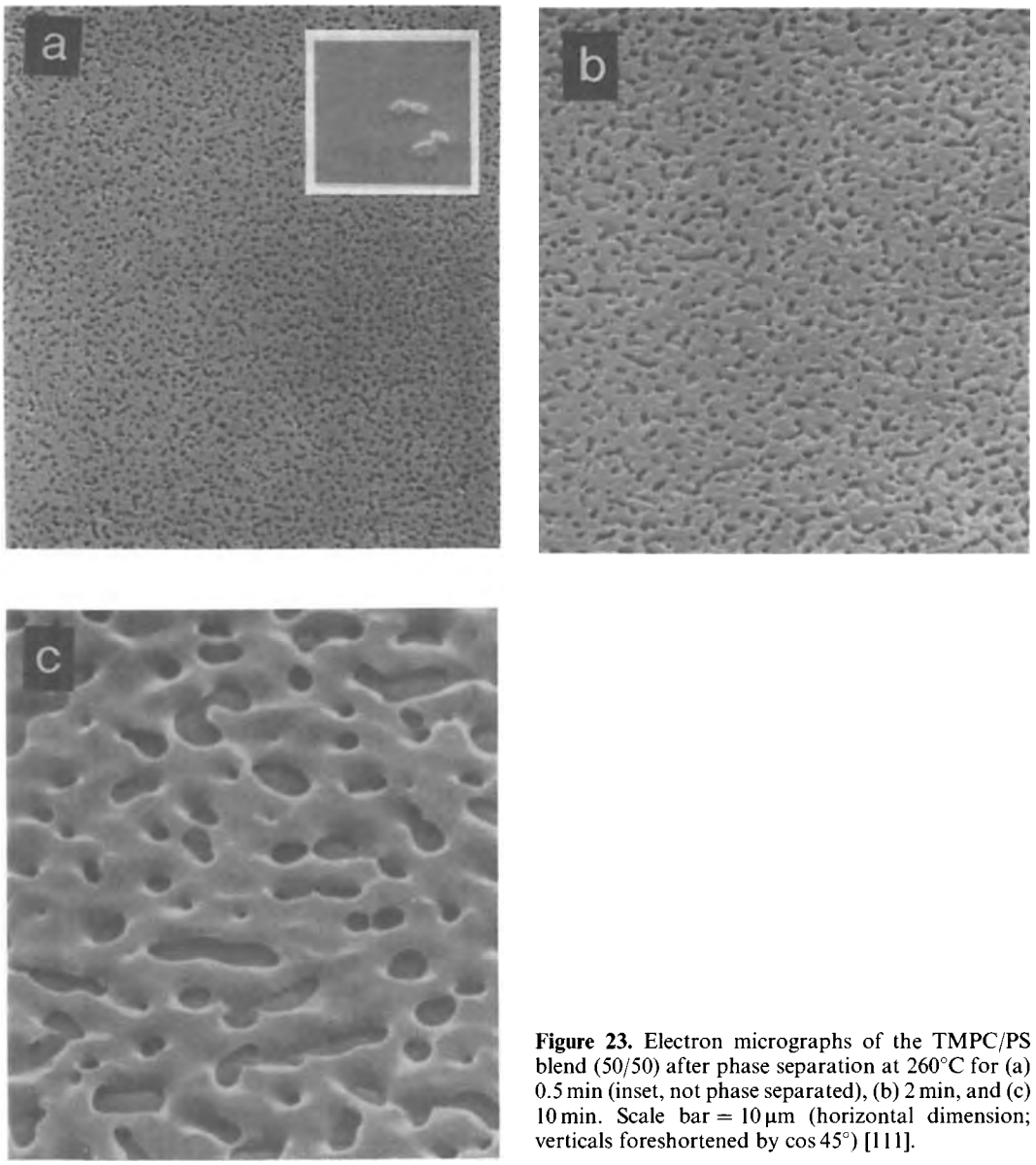


Figure 23. Electron micrographs of the TMPC/PS blend (50/50) after phase separation at 260°C for (a) 0.5 min (inset, not phase separated), (b) 2 min, and (c) 10 min. Scale bar = 10 μm (horizontal dimension; verticals foreshortened by cos 45°) [111].

die drawing [122], and superdrawing [123]. In addition to polyethylene, high modules fibers of polyacetal [124], polypropylene, polyamide hydrazides [124], and others are known.

Electron microscopic investigation of such fibers revealed deformed crystalline lamellae perpendicular to the draw direction, separated by highly drawn fibrils parallel to the draw direction in polyethylene [125]. In later work, details of the internal structure of such fibers were revealed by permanganic etching [126], using conventional transmission electron

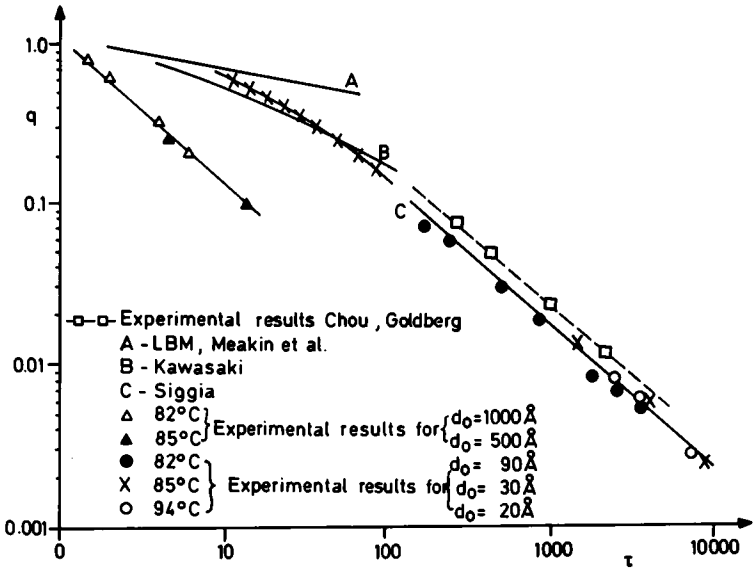


Figure 24. Inverse size of phase-separated regions versus time (dimensionless units) [113].

microscopy and by scanning transmission electron microscopy (STEM) [127]. It was found that deformation is initially accommodated by the inter lamellar regions, which cavitate and form microfibrils. With increased deformation, suitably oriented lamellar crystals deform by two chain-slip systems: {100}, ⟨001⟩; and {010}, ⟨001⟩. These *c*-axis shear processes (along the molecular axis) were further distinguished as fine slip or block shear (Fig. 25). When the block size is below a critical value, the blocks form highly oriented fibers with discrete microfibrils containing small block crystals and strain-induced fibrillar crystals (Fig. 26).

13.2.4.2 Liquid-Crystal Fibers

Stiff, rod-like molecules which exhibit nematic liquid crystalline behavior in the melt (thermotropic) or in solution (lyotropic) can be transformed into highly oriented extended chain structures by self-alignment. Thousands of liquid crystalline polymers are documented in the literature [128], the most important of which are:

- poly(*p*-phenyleneterephthalamide) produced by Du Pont under the tradename Kevlar;
- aromatic copolyesters produced by Hoechst-Celanese under the tradename "Vectra";
- a group of lyotropic, aromatic heterocyclic polymers which exhibit the highest fiber modulus yet observed [129, 130].

In a detailed study elucidating the structure–property relations of liquid crystalline polymer fibers, SEM revealed a microstructure composed of fibrillar units organized in a hierarchical structure (Fig. 27) [131]. The mechanical properties of these materials are related to the stress transfer between the individual structural units. It is generally accepted that full, three-dimensional crystallographic

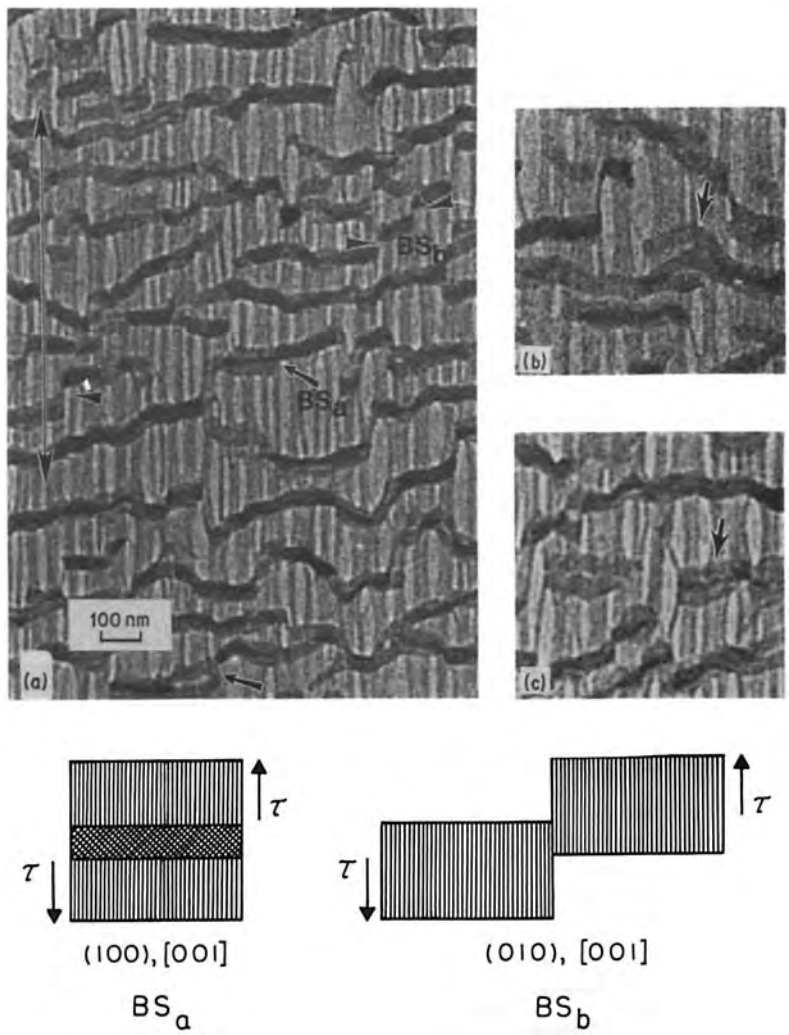


Figure 25. (a) Conventional TEM bright-field micrographs of deformed AN films. BS_a and BS_b refer to lamellae deformed by block shear. (b) Lamellae (arrow) sheared by combined block shear (100) and fine shear (101). (c) Lamellae (arrow) having undergone block shear (100) producing overlapping lamellae. (d) Schematics of observed block shear processes [127].

registry is not necessary for efficient stress transfer [132].

13.2.5 Liquid-Crystalline Polymers

The fascinating physical properties induced by the self-ordering of large molecules with asymmetrical shapes (mesogens) is a well-known phenomenon [133, 134] and has been studied extensively in large monomeric molecules, which are generally synthesized with specific functionalized groups of atoms.

Liquid crystalline polymers are a comparatively new class of materials, which

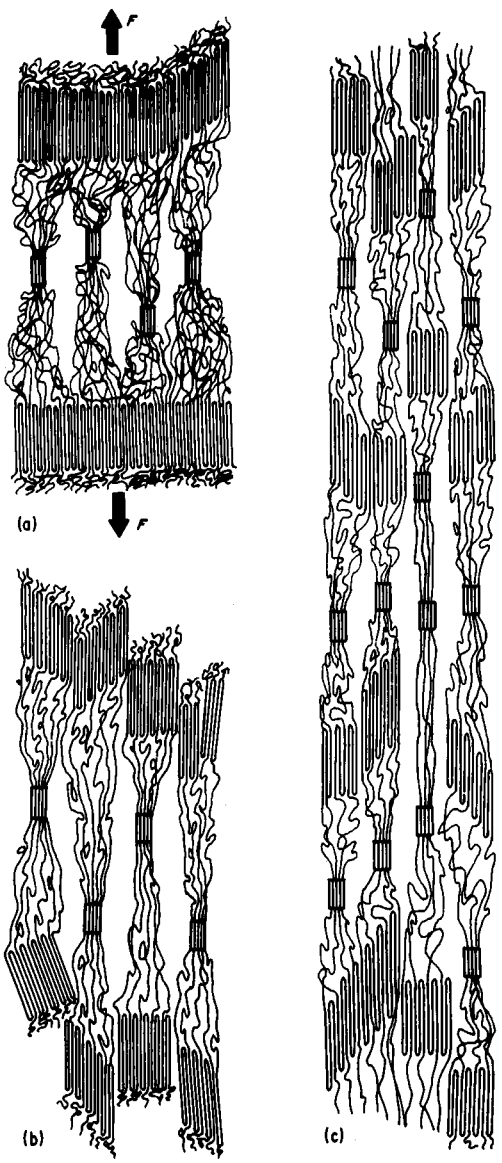


Figure 26. Ultrastructure of deformation: (a) strain in amorphous regions sufficient to induce crystallization in the fibers; (b) block slip resulting in break up of lamellae into large blocks; (c) fibrillar morphology with discrete microfibers containing some lamellar block crystals and strain induced fibrillar crystals [127].

combine the useful optical, electrical, and ferroelectric properties of liquid crystalline monomers with those of mechanical strength and considerably higher transition temperatures. In these materials the mesogens are generally attached to a polymer chain by alkane spacers, so that the materials retain their mobility and self-ordering characteristics without being too severely impeded by the tendency of the polymeric chain to form a Gaussian coil [1, 134, 135]. The thermodynamic properties of rod-like molecules were first treated by Flory [136] and the theory later refined by Saupe [137], while those of tilted layers and stacks of discs were discussed by McMillan [138].

Typical mesogenic geometries are shown in Fig. 28. The molecular shape and manner of attachment to the polymeric chain influence the nature of the mesophase, and the temperature interval in which it is formed, thus producing many useful materials (Fig. 29). In discussing their structural characteristics it is important to note that these materials possess only one- or two-dimensional quasi-long-range order, not three-dimensional long-range order as crystals do. Therefore they should be described by their pair correlation functions $g(2)$. The relationship between the pair correlations obtained for liquids, liquid crystals, and crystals is shown in Fig. 30a. The effect of the lattice statistics in layer structures on the structure factor $S(q)$ is shown in Fig. 30b. Although most structural investigations on liquid crystals have been performed using X-rays or neutrons, the structural information that can be obtained in reciprocal space is meager: Broad, oriented halos, in the wide angle range, loss of higher order small angle reflections due

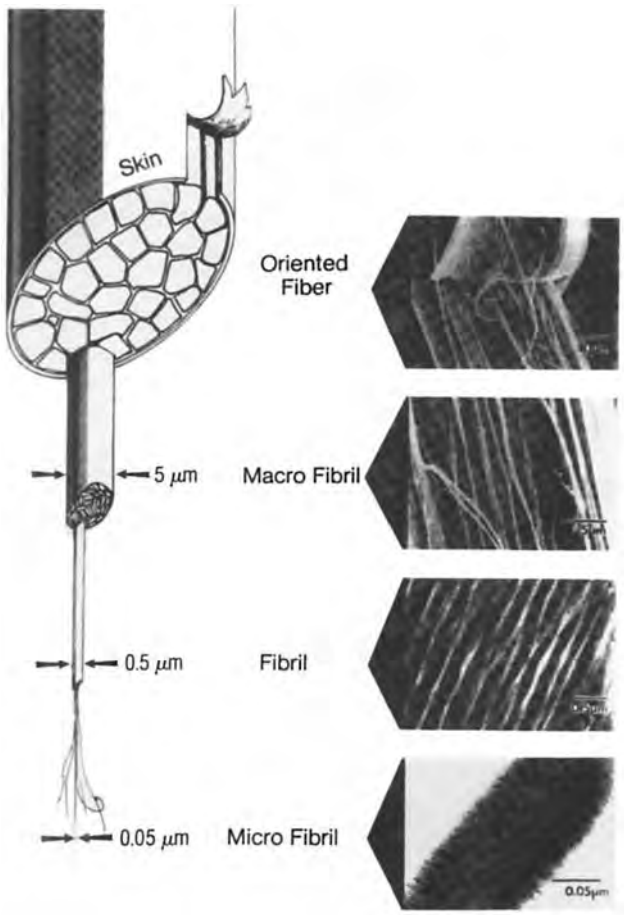


Figure 27. The long-chain polymer structural model shows the hierarchical, fibrillar texture of long-chain polymer materials in an artist's drawing (left) and in micrographs of oriented materials (right) [131].

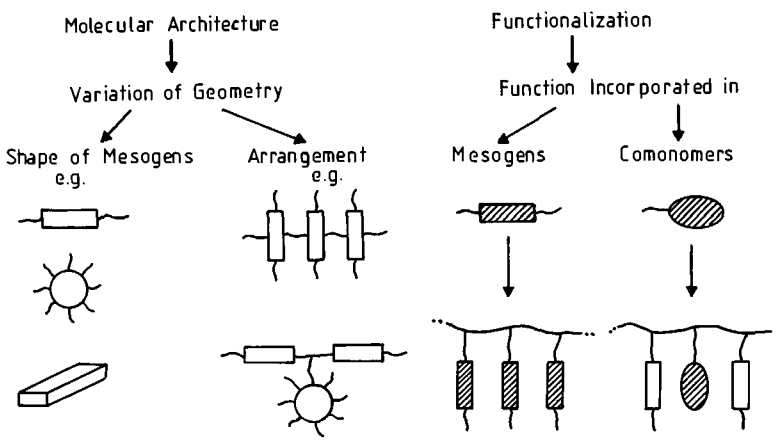


Figure 28. Molecular engineering of long-chain polymers.

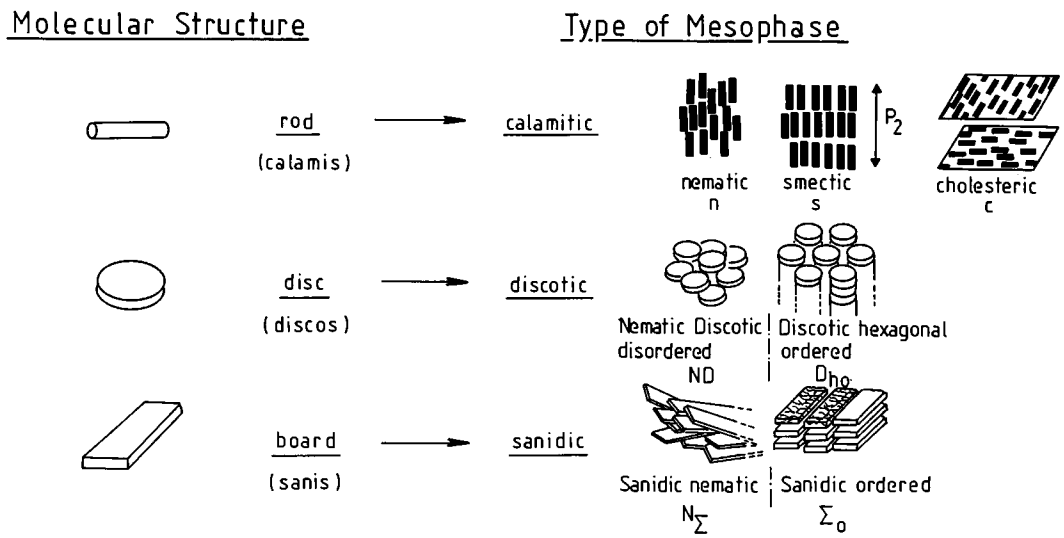


Figure 29. Effect of molecular shape on type of mesophase formation.

to the loss of short-range correlations, and loss of mixed reflections due to the lack of three-dimensional order. For this reason a real-space method such as electron microscopy is much more informative. The difficulty lies, as so often in electron microscopy, in finding a suitable method of sample preparation, image recording, and image interpretation, and in understanding the relation between the structure and the physical properties under investigation. As is usual for polymers, each material requires a special method of preparation and analysis. Therefore electron microscopic investigations on this fascinating class of materials are relatively limited. It is hoped that this chapter will stimulate such research.

Initial low resolution TEM investigations were concerned with the study of the

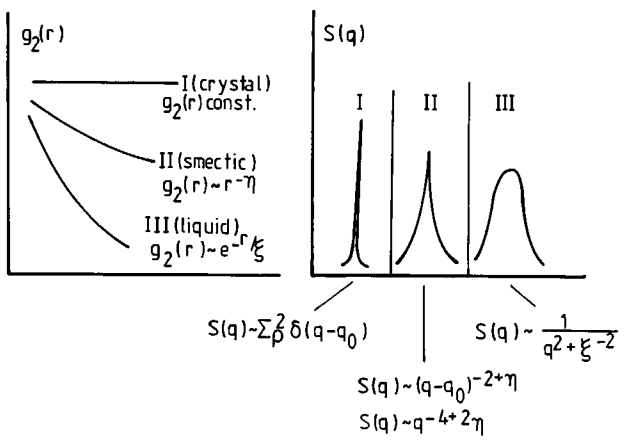


Figure 30. Order in smectic liquid crystals.

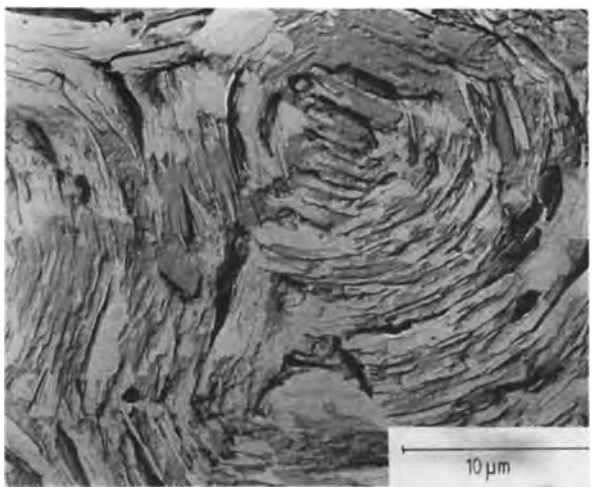


Figure 31. Image of $s = \pm 1$ disclination in ferroelectric chlorobiphenyl.

characteristic walls in thermotropic nematic copolyesters [139–141]. In a later paper on a similar system, two types of defects were observed: Néel bend walls, and $s = \pm\frac{1}{2}$ disclination lines [142]. Subsequently, other authors found $s = \pm 1$ disclination loops in a ferroelectric chlorobiphenyl [64] (Fig. 31) and parabolic focal conics in a semiflexible liquid crystalline polyester [143].

In order to interpret images of liquid crystalline polymers correctly some fundamental properties of liquid crystals need to be discussed. Liquid crystalline mesophases can be divided into four general classes, which are related directly to the shape of the mesogens. The defects are specific for these classes and are related to the elastic energy density of the system caused by bend, twist, and splay of the molecular layers.

Nematic Phase

The mesogens are arranged in such a way that their centers of mass display short-range order and their long axes, described by director $\bar{n}$, lie preferentially in one direction, producing long-range orientational order. Mesogens with twisted nematic phases are typically used as display devices. In the light microscope, the nematic phase displays typical defect structures [144] resulting from the three principal types of elastic distortion. The elastic constants K_{11}, K_{12}, and K_{33} correspond to splay, twist, and bend, respectively. The free energy density of the deformed system is given by [145]

$$F = \tfrac{1}{2}[K_{11}(\nabla \cdot n)^2 + K_{22}(n \cdot \nabla \times n)^2 + K_{33}(n \times \nabla \times n)^2]$$

The total energy of the system is:

$$E = \int F \, dr$$

Minimization yields a set of differential equations which describe the director field. If the director is confined to the xy plane,

$$n_x = \cos\phi, \; n_y = \sin\phi, \; n_z = 0$$

$$\phi = s\alpha + c$$

where $\alpha = \tan^{-1}(y/x)$, c is a constant, $s = \pm\frac{1}{2}, \pm 1, \pm\frac{3}{2}$, and $0 < c < \pi$ [146].

The director field for nematics according to this expression can be simulated and the corresponding structures in polymeric liquid crystals investigated by electron microscopy. The agreement between calculated and observed structures (for $s = \pm\frac{1}{2}$) was elegantly demonstrated thus by Thomas [63].

Smectic Mesophase

Rod-shaped mesogens are arranged in layers, so that the centers of mass now have quasi-long-range positional order. Within the layers their long axes are preferentially ordered in one direction. Mobility is possible about the long molecular axes and between layers. Frequently, benzoic acid phenylesters are used as mesogens. More recently, biphenyls and phenylcyclohexanes have been used. A comprehensive summary is given by Finkelmann [147]. Frequently, the mesogenic groups have been functionalized by dye-containing groups, by groups that undergo photoisomerizations, or chiral groups which induce tilt and destroy centrosymmetry [135]. Eleven types of smectic mesophase have been postulated; they differ in the order of the mesogens within the layers, ranging from smectic A (no order within layers, molecular long axis perpendicular to layers) via smectic C (no order within layers, molecular long axis tilted with respect to the layers), to various degrees of order within the layers (hexagonal, orthorhombic, cubic) with long axes either perpendicular or tilted with respect to the layers. The interest in layer structures lies in their ability to form sheets of highly directed functionalized groups by the application of an electric or magnetic field in the liquid-like mesophase. This may be smectic A for second-harmonic generation (the frequency of incoming laser light is doubled during transmission) [148, 149] or smectic C for ferroelectric applications (the direction of the dipole is reversed when the applied electric field is reversed) [150]. It is evident that structural defects play a major role in the applicability of such materials and that electron microscopy is an ideal tool to study them. As an example, consider the study of ferroelectric molecules for applications as display devices.

The elastic energy density in tilted smectic phases which are capable of giving a ferroelectric response is given by [146]

$$F = \frac{1}{2}\left\{\left(\frac{\partial\varphi}{\partial x}\right)^2 + \left(\frac{\partial\varphi}{\partial y}\right)^2 + \left[\left(\frac{\partial\varphi}{\partial z}\right) - q\right]^2\right\}$$

where the director C spirals around the C axis (perpendicular to the layers) such that:

$$\varphi = qz \quad \text{and} \quad q = \frac{2\pi}{p}$$

Minimization such that $\nabla^2\varphi = 0$ yields wedge dislocations along the z axis with

$$\varphi = s\tan^{-1} y$$

and twist dislocations along the y axis with

$$\varphi = s\tan^{-1}\left(\frac{z}{x}\right) + qz$$

where s is a constant which is ± 1 for smectic C meso phases.

These expressions were calculated analytically to give the director fields shown in Fig. 32. Comparing these with electron micrographs from suitably prepared samples shows the excellent agreement, and identifies the $s = \pm 1$ disclinations in Fig. 31 [64]. As well as the structural defects, a further parameter dictates the applicability of these molecules for display purposes,

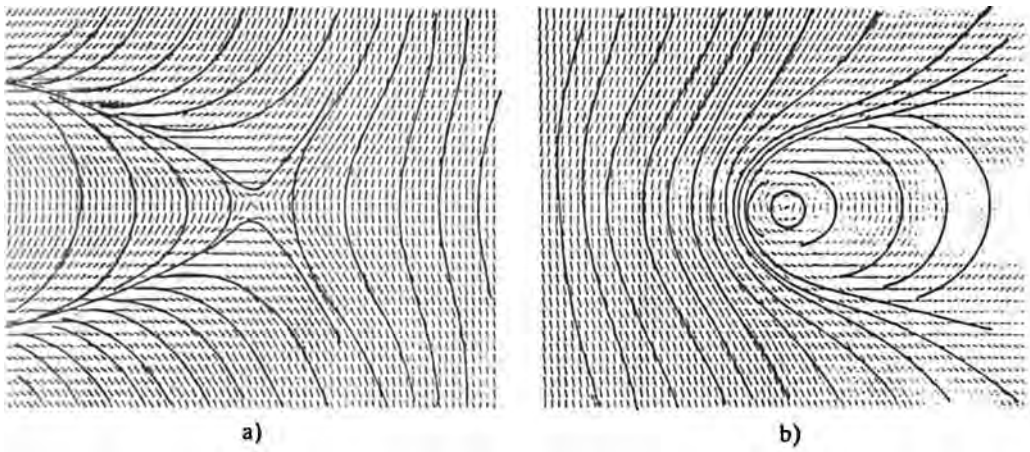


Figure 32. Calculated director field for $s = \pm 1$: (a) $s = 1$, $q = 0.1$; (b) $s = -1$, $q = 0.1$ [64].

namely their switching time. In this respect, the monomeric analogs are undoubtedly superior to the polymers. On the other hand, the polymers give stable, solid films at room temperature with highly directed dipoles.

In addition to these rather large-scale defects, it became clear that microscopic details regarding the layer structure were required for an understanding of the translational and orientational correlations in both smectic A and smectic C structures. Electron diffraction patterns indicated that the relevant features lie in the small angle region corresponding to 10–20 Å in real space. Therefore it is necessary that these spatial frequencies are transferred in the electron microscope for imaging. It was proposed that a rather unusual microscope transfer function χ must be used, in order to have $\sin \chi = 1$ in the appropriate range of spatial frequencies [151]. On the basis of these considerations, defects were revealed by high resolution electron microscopy and shown to play a decisive role in smectic systems. These defects are undulations and disclinations (Fig. 33) [152–154].

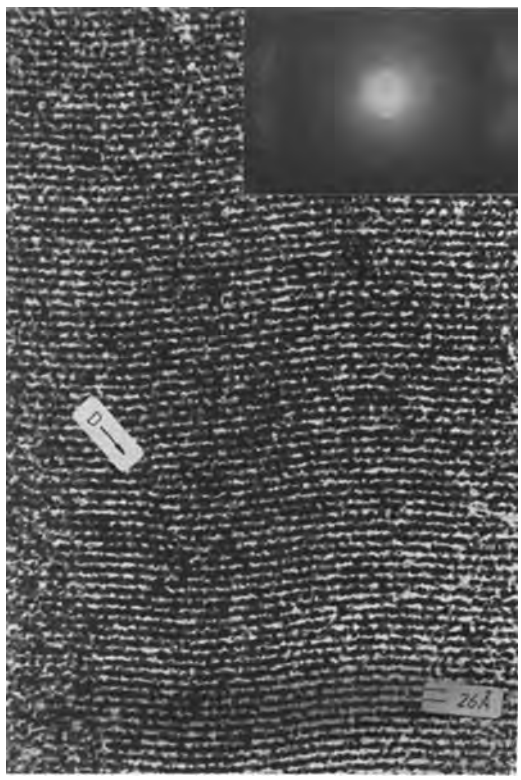


Figure 33. High resolution electron micrograph of a main chain/side group liquid crystalline polymalonate [153].

The observed undulations show that smectics do not conform to our usual notions of long-range order. In fact, long-range order is impossible in one dimension, as was shown by Landau and Peierls [155] because the mean square fluctuations $\langle u^2(r)\rangle$ are divergent:

$$\langle u^2(r)\rangle = (4\pi)^{-1}(BK)^{-\frac{1}{2}}kT\ln\left(\frac{q_{\max}L}{2\pi}\right)$$

where q is the scattering vector at which the maximum intensity is observed, T is the temperature, and B and K are the elastic constants for compression and curvature, respectively. The quasi-long-range order in smectic liquid crystals has been studied extensively by X-ray scattering in fundamental work by Als-Nielsen et al. [156]. From high resolution images more detailed information about the precise nature of this quasi-long-range translational order was obtained and related to the phenomena observed in the electron diffraction patterns [151–153]. The comparison revealed the expected sharp diffraction maxima, with only minimal diffuse scattering in the wings of the peaks. In addition, characteristic streaks on the diffraction maxima indicated the loss of long-range translational correlations in one direction. With the aid of high resolution images, the origin of the diffraction phenomena could be identified uniquely as small-amplitude (about 10 Å) undulations of long wavelength (several micrometers) [153, 157].

Discotic Mesophases

Discotic mesophases are formed from disc-shaped mesogens arranged in columns. The discs may be ordered or disordered within the columns, perpendicular or tilted with respect to the columns, and the columns may be arranged in a hexagonal, orthorhombic, or quadratic manner. Mobility is possible between the columns as well as within the columns (e.g. rotation of the discs). Typical examples are the triphenylene esters and ethers. The lattice statistics for these two-dimensional systems were derived by Selinger and Bruinsma [158] and one expects quasi-long-range order, experimentally indicated by δ functions for the diffraction peaks at each reciprocal lattice vector, surrounded by diffuse thermal scattering in the wings.

Such sharp, small angle maxima were indeed observed in an electron microscopic investigation on discotic polymers [159]. In addition, the high resolution images were evaluated by cross-correlation methods and the deviation vectors gave detailed information about the precise direction and size of deviations from perfect lattice positions [159, 160] (Fig. 34). In addition, the detailed structural analysis of monomeric analogs from electron diffraction patterns has given precise information about atomic positions; this is described in the chapter on electron crystallography (see Sec. 4.1.2 of this Chapter).

Sanidic Liquid Crystals

Sanidic, or 'board-like' liquid crystals are fully aromatic polymeric amides/esters/enamines with alkyl side-chains. These stiff molecules were originally synthesized in order to produce a high-tensile-strength material, but the material was found to be insoluble; to produce solubility and to reduce the melting temperature, flexible spacers, kinks, or side-groups were added to the polymer backbone. Both low level substituted polymers [161] and highly

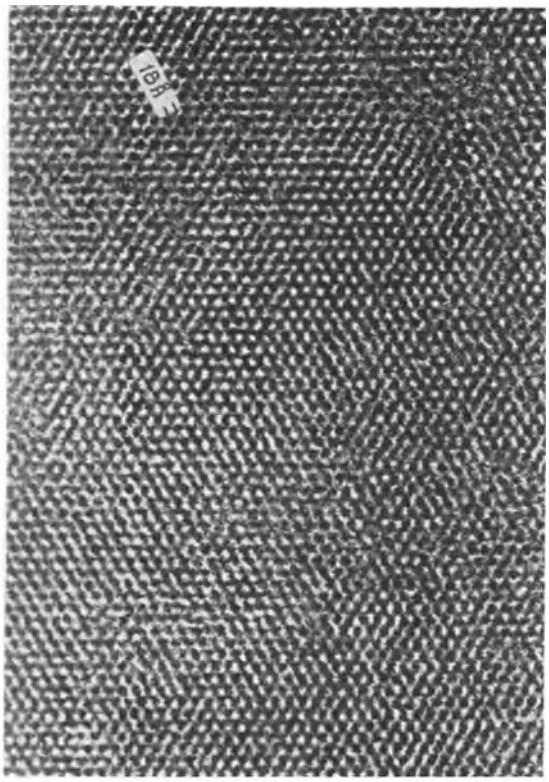


Figure 34. (a) Columns of discotic triphenylene molecules viewed from the top [159]. (b) Deviation vectors obtained by cross-correlation of discotic 'lattice' with perfect crystal [159].

substituted systems [162] were investigated by X-ray diffraction. Because of the appearance of two perpendicular rows of diffraction maxima, the board-like structure was proposed. The dramatic loss of tensile strength in these materials was not understood. In a detailed investigation by high resolution electron microscopy, the loss of mechanical strength could be related to the lack of coherence between small, rather poorly oriented domains (Fig. 35) [163]. Detailed structural analysis by electron diffraction of monomeric analogs revealed that the molecule cannot adopt a flat board-like conformation [164].

13.2.6 Organic Molecules with Functional Groups

A major aim in organic materials science is to design a molecular architecture that satisfies specific criteria with regard to the desired physical properties. In order to achieve this it is essential to know the molecular conformation and packing in the condensed state, as well as details regarding the superstructure and defects.

In recent years organic polymeric molecules have been synthesized in order to produce one of the following properties: ferroelectricity, photoconductivity, mechanical strength, or a second-order NLO effect. In order to achieve these properties, it is necessary to satisfy specific structural criteria as described in the following.

13.2.6.1 Ferroelectricity

Rod-shaped molecules must pack in such a way that their structure is

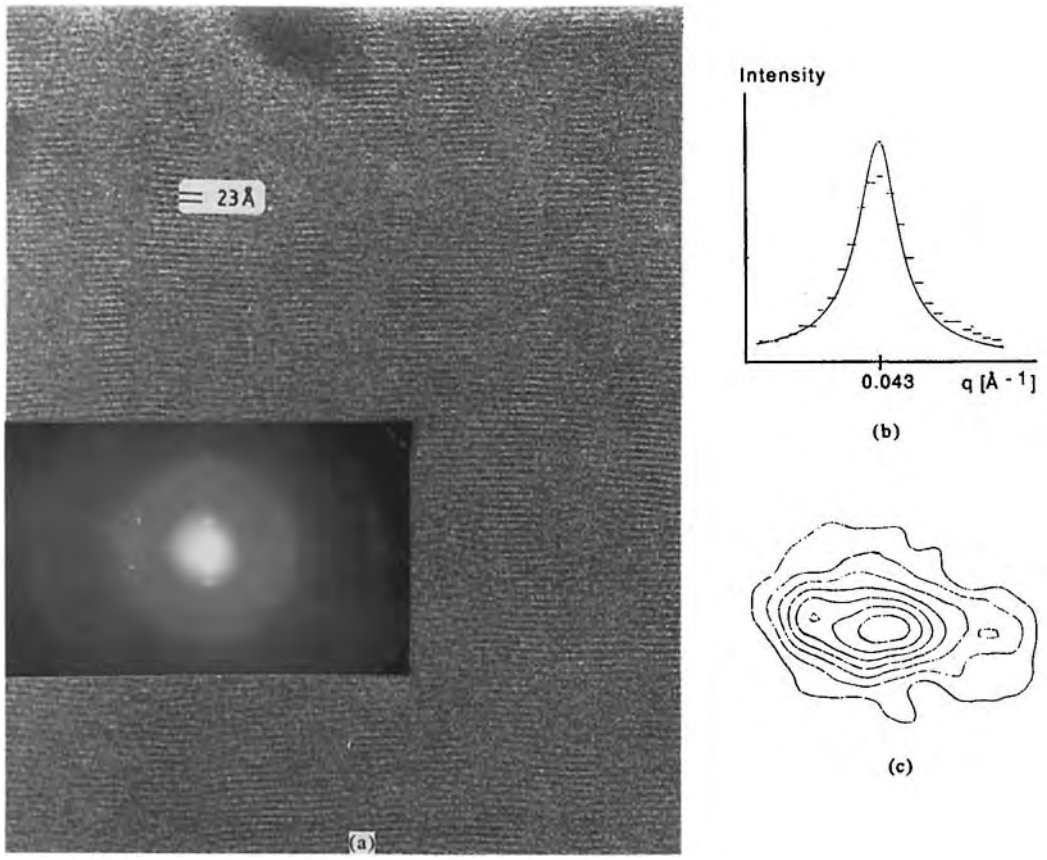


Figure 35. Line profile analysis of diffraction spots from poly(1,4-phenyleneterephthalamide) derivative: (a) micrograph and diffraction pattern; (b) intensity distribution (Lorentz fit); (c) spot-contour lines of second-order peak with satellites [163].

noncentrosymmetric and the dipoles μ are oriented so that there is a resultant polarization P in a direction which can be reversed on application of a field. The molecular dipole μ and the macroscopic polarization P are related by the formula:

$$P \propto \sum_i \langle \mu_i \rangle N_i$$

where N_i is the number density of dipoles. A ferroelectric effect can be induced if the molecule has a molecular subgroup with a suitably oriented dipole μ plus one or two chiral groups to induce molecular tilt and to prevent the formation of a centrosymmetric superstructure. The chiral group induces a helical superstructure which must first be eliminated by suitable alignment, thus giving rise to a tilted layer structure with a macroscopic polarization P.

Switching occurs if the molecules can be rotated collectively within the smectic layers by application of an electric field $\boldsymbol{E}$ [64]. In order to optimize the properties, it is clearly essential to understand the relationship between molecular conformation and dipole moment as well as the direction

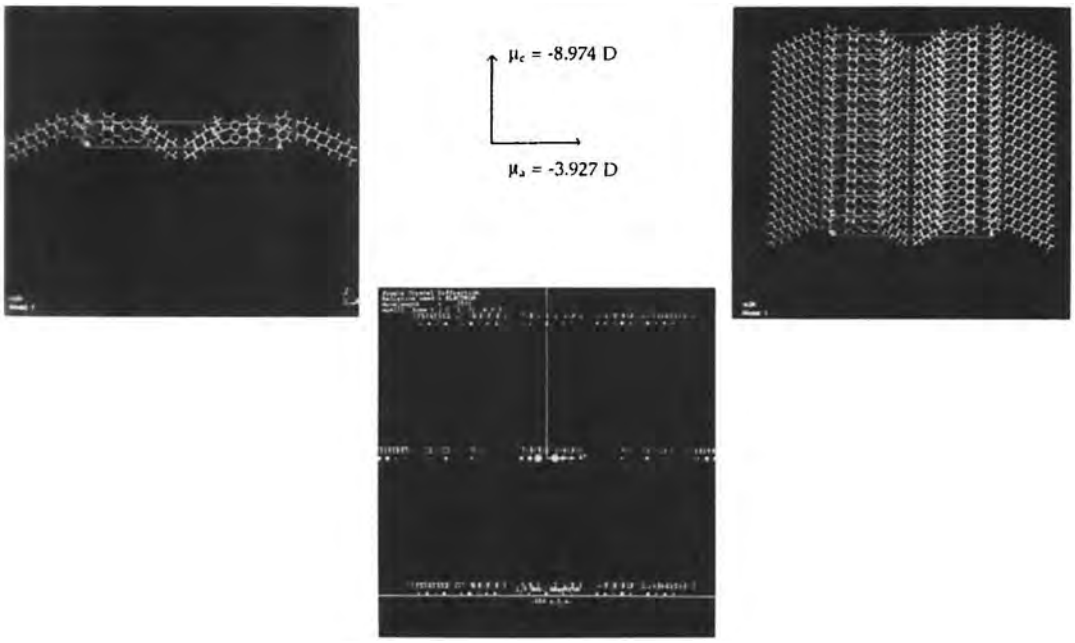


Figure 36. Unit cell, crystal, and diffraction pattern of a ferroelectric molecule, showing the direction of the resultant dipole [165].

of the dipole within the layers [165]. Typical molecules belonging to this category are biphenyl ester and thiadiazol compounds. In order to optimize the effect, the molecular dipole μ must have a large value perpendicular to the molecular layers. If molecules are to be synthesized with this aim in mind, their packing principles must be fully understood and the molecules oriented so as to produce the desired resultant polarizability P. It has been shown by electron diffraction combined with high resolution imaging that this information can be obtained from the monomeric analogs because these provide a sufficient number of reflections to perform structural analysis and derive the direction of the dipoles (Fig. 36). Although the polymers provide insufficient diffraction maxima for ab initio analysis, the relationship between intensities available from the polymer and those of the monomer can be established [165] and structural information about the polymer derived from the monomer.

13.2.6.2 Photoconductivity

Charge transfer, expressed in terms of the mobility of charge carriers, is possible perpendicular to thin organic films if, for example, suitable disc-shaped molecules are oriented in stacks perpendicular to the films [166]. In order to induce photoconducting properties, it is necessary to optimize the relationship between the molecular conformation, their mutual packing, as well as superstructural ordering. Typical examples are hexakis(heptyloxy)triphenylenes [159, 160]. The effective mobility m perpendicular to the layers is

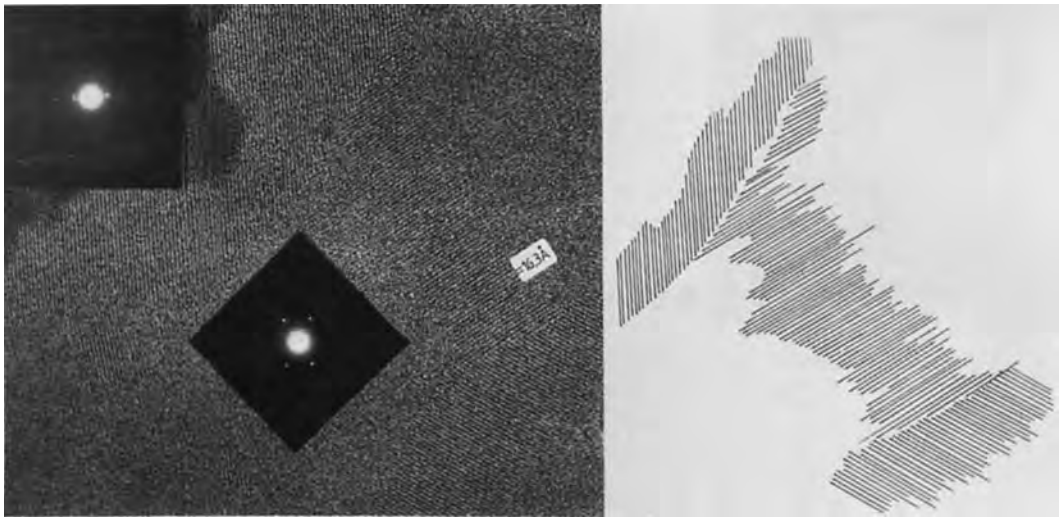


Figure 37. (Left) high resolution electron micrograph of crystalline triphenylene monomer with microdiffraction pattern in correct orientation. (Right) transparent overlay showing a grain boundary [159].

given by:

$$m = \frac{l^2}{t_T V}$$

where t_T is the transit time, l is the sample thickness, and V is the applied voltage across the sample. Transport occurs across the triphenylene cores and is affected by the distance between them, detailed structural parameters are required which can be obtained from electron diffraction data. In addition, information about superstructural parameters such as grain boundaries which act as local traps can be obtained from electron microscopic images.

From electron diffraction experiments combined with molecular modelling the structural space groups and molecular orientation were calculated. In the triphenylene ethers it was found that in the nonconducting crystalline phase, the cores of the ethers were arranged with orthorhombic symmetry and were tilted with respect to the layers [159], while those of the esters were arranged hexagonally and untilted [160]. In both cases the mutual stacking of the columns with respect to each other was such that a superstructure was created as well as numerous grain boundaries (Fig. 37). During the transition to the conducting discotic mesophase, molecular tilt vanishes, and all superstructure reflections are lost due to the onset of disc rotation while, however, retaining the average column positions and eliminating grain boundaries [165]. Crystal grain boundaries are eliminated in the discotic phase (Fig. 34a), so that charge transfer is not inhibited by local traps as in the crystalline phase.

13.2.6.3 Mechanical Strength

Sanidic, or board-like, molecules are fully aromatic polymeric amides or esters with alkyl side-chains. Such materials were expected to have good mechanical

properties because of the stiff polymeric backbone. In practice, the materials fell far short of expectation. The explanation was found to be related to details of molecular packing [163, 164]. The oriented regions were shown to be highly imperfect and of small dimensions. Using semiempirical quantum mechanical calculations (see Sec. 4.2.1 of this Chapter) the aliphatic chains at the terminal phenyl groups were shown to be rotated out of the plane containing the central phenyl group. The molecular conformation therefore deviates from the previously proposed board-like structure, because the aliphatic chains project out of the plane containing the phenylene core.

The molecular packing in the condensed phase was then determined by simulation of the electron diffraction patterns from the monomeric analogs in three different projections. The model structure obtained from electron diffraction (Fig. 38) immediately explains the previously puzzling observation that X-ray small angle scattering experiments always gave long spacing considerably below the extended chain length. This is due to the screw-like arrangement of the side-chains around the molecular backbone. The high resolution images indicated that the loss of mechanical strength arises because oriented regions are very small and poorly developed (Fig. 35).

13.2.6.4 Second-Harmonic Generation

The nonlinear optical properties of individual molecules are related to their transition moment β. In microscopic terms, the induced dipole moment μ in an electric field $\boldsymbol{E}$ is given by:

$$\mu_i = \mu_i^0 + \alpha_{ij}\boldsymbol{E} + \beta_{ijk}\boldsymbol{E}^2 + \gamma_{ijkl}\boldsymbol{E}^3 + \cdots$$

where μ_i^0 is the permanent dipole moment, α_{ij} is the molecular polarizability, β_{ijk} is the second-order hyperpolarizability, γ_{ijkl} is the third-order hyperpolarizability, and i, j, and k are the molecular axes.

The hyperpolarizability β_{ijk} of the molecule is related to the macroscopic second-order susceptibility χ_{IJK} of the crystal by the relationship:

$$\chi_{IJK}^{(2)}(-2\omega;\ \omega,\omega)$$

$$= \frac{1}{V} f_i^{(2\omega)}(\omega) f_j^{(2\omega)} f_k^{(2\omega)} \mathbf{R}\beta_{ijk}$$

where $f(\omega)$ are orientation distributions, $\mathbf{R}$ is the transformation matrix from molecular (i,j,k) to crystallographic (I,J,K) coordinates.

In a macroscopic experiment, the following parameters are determined:

$$\boldsymbol{P}_I = \chi_{IJ}^{(1)}\boldsymbol{E}_J + \chi_{IJK}^{(2)}\boldsymbol{E}_J\boldsymbol{E}_K + \chi_{IJKL}^{(3)}\boldsymbol{E}_J\boldsymbol{E}_K\boldsymbol{E}_L$$

where $\boldsymbol{P}_I$ is the polarization in the medium. Therefore, in order to obtain a second-order susceptibility χ_{IJK}, the structure must be noncentrosymmetric and the major component of the molecular transition moment β_{ijk} must be oriented in the desired macroscopic direction, namely in the plane of the oriented film. It is clear that a contribution can be made toward developing and synthesizing suitable materials with the desired properties only if the packing principles of the molecules (i.e. their conformation and orientation with respect to the macroscopic axes) are understood.

In an initial structural investigation on a dione, the direction of the molecule and its dipole in the unit cell were determined

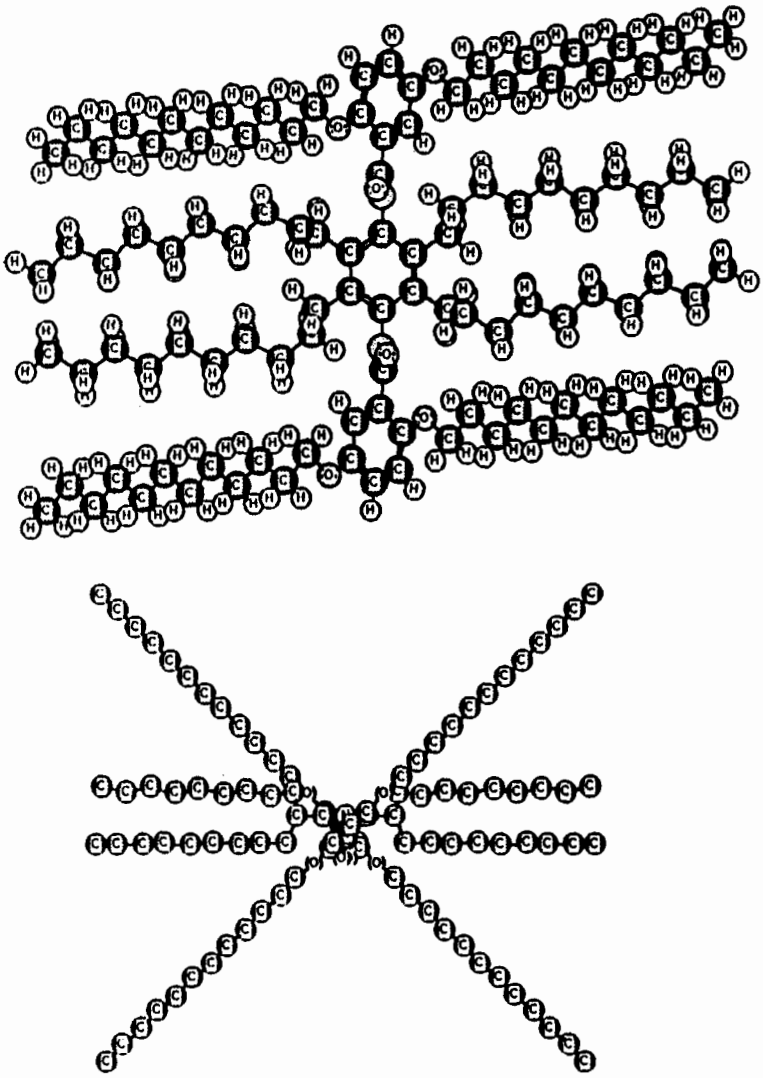


Figure 38. Molecular conformation of a substituted molecule in the crystalline phase from two different projections [164].

(Fig. 39) [149]. In a more detailed calculation, the conformation and orientation of the molecular head group were determined using both simulation methods and maximum entropy statistics [167]. The electron microscopic investigations indicated that the dione molecule in the required conformation has a good value of β but crystallizes in a centrosymmetric $P\bar{1}$ space group, thus destroying its ability to produce a second harmonic response. However, if the same head group is attached to an aliphatic side-chain, centrosymmetry is destroyed, but molecules in the unit cell are arranged so that most of the major components of β are mutually cancelled, leaving only a small resultant effect (Fig. 39). Finally, however, the molecular arrangement in Langumir Blodgett films avoid mutual cancellation and

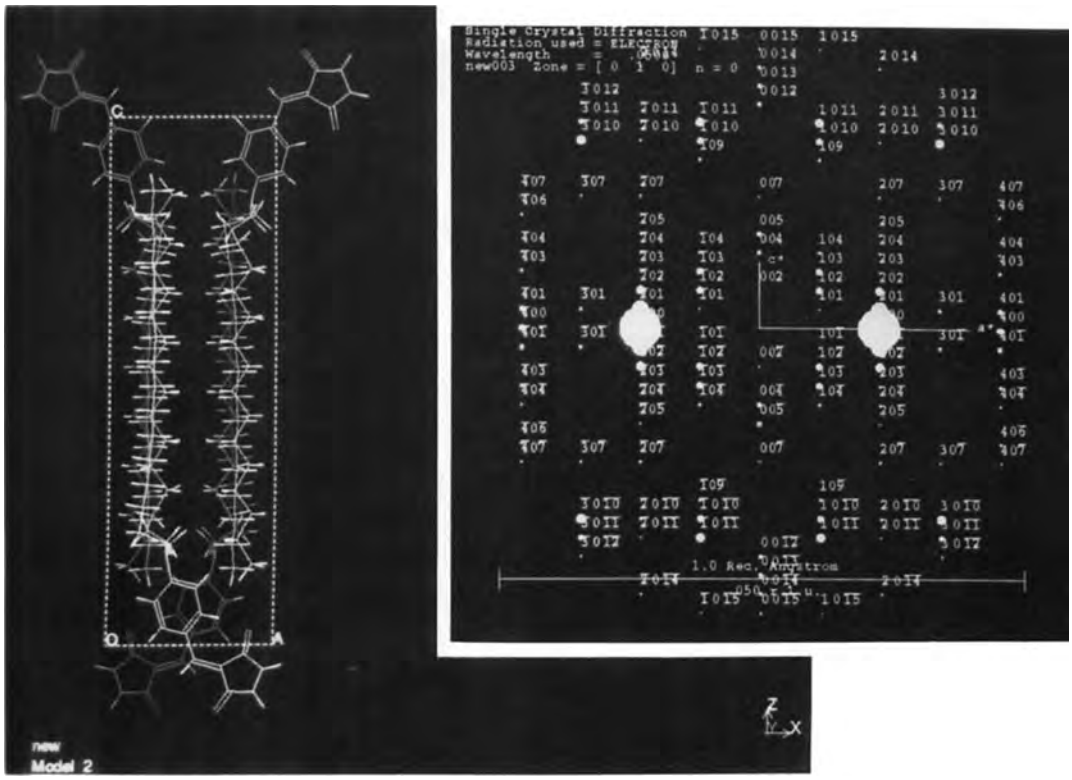


Figure 39. Orientation of a dione molecule in the unit cell, giving a resultant hyperpolarizability per unit cell of $\beta_c = 0.765 \times 10^{-30}$ e.s.u.

centrosymmetry, giving rise to the maximum value of χ, as expected when the molecular origins are understood [149, 168].

13.3 The Use of New Microscopy Methods to Study Polymers

13.3.1 Atomic Force Microscopy

Extensive use of atomic force microscopy (AFM) has been made in the study of polypropylene by Lotz et al. [169]. The situation for the investigations was particularly fortunate because the investigators had already performed detailed TEM investigations on this molecule and were thus able to use prior knowledge and experience as well as to being in a position to compare the results obtained with the two methods. From the TEM investigations of syndiotactic polypropylene the authors had already proposed a unit cell with $a = 14.5$, $b = 11.20$, and $c = 7.4$ Å and a space group *Ibca* [170] resulting from the presence of both right- and left-handed helices in strict alternation (Fig. 40). In addition, at lower crystallization temperatures, crystals having a unit cell with $a = 14.5$, $b = 5.6$, and $c = 7.4$ Å

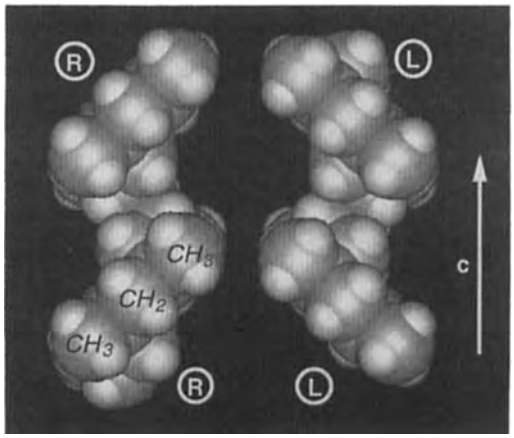


Figure 40. Computer-generated molecular models of right- and left-handed sPP helices in the $(t_2g_2)_2$ conformation, as seen along the crystallographic *a* axis (left- and right-hand sides of the figure, respectively). Note the prominent rows of CH_3, CH_2, and CH_3 groups at 45° to the chain-axis direction, which reveal unambiguously the helical handedness [175].

and a space group $C222_1$ are formed. In this case, the helices all have the same hand. In actual crystals, both packing schemes were found to coexist, where shifts of $b/4$ produced by 'mistakes' in the alternation of helical handedness in the *bc* planes were proposed. These cause streaking on $(hl0)$ reflections parallel to a^*. Summarizing these investigations, it can be said that they made a considerable contribution towards understanding the general principles of helical packing in polymeric molecules.

When applied to suitable polymers, AFM has reached methyl group resolution, as illustrated by the visualization of end groups and fold structures in crystals of linear and cyclic paraffins [171] of chain structure in thin films of PTFE [172, 173], poly(oxymethylene) [174], and isotactic polypropylene [170]. Therefore it was a particularly well-suited method to investigate the defects in helical packing on the *bc* planes of syndiotactic polypropylene [175]. In order to expose this face, a complicated series of preparative steps were necessary, which are described fully in the reference cited. In addition, AFM examination itself proved to be a challenge because of the softness of the polymer film and adhesion of the tip to the surface. AFM imaging in a liquid environment gave much improved resolution because a reduction of capillary forces made it possible to use imaging forces in the region of 10^{-10} N and below. It emerged that AFM recognition of the helical hand in synthetic polymers is more difficult than for larger double strands like DNA because the helices are not stabilized by intermolecular forces. The final results, shown in Fig. 41, confirm the striking regularity of the helical hand alternation with a crystal structure and symmetry consistent with the *Ibca* space group, after Fourier filtering of the unfiltered images. In a series of studies, the same authors examined the contact faces of epitaxially crystallized α- and γ-phase isotactic polypropylene [176].

13.3.2 Scanning Tunneling Microscopy

The structure of organic molecules and polymers at the surfaces of specific substrates has been investigated successfully by scanning tunneling microscopy (STM) [176, 177]. In order to achieve images, the material needs to be able to transport electrons from the emitter tip to a conductive support, either by electron tunneling for thin samples of about 50 Å, or by an additional conductive mechanism for thicker samples. Most successful

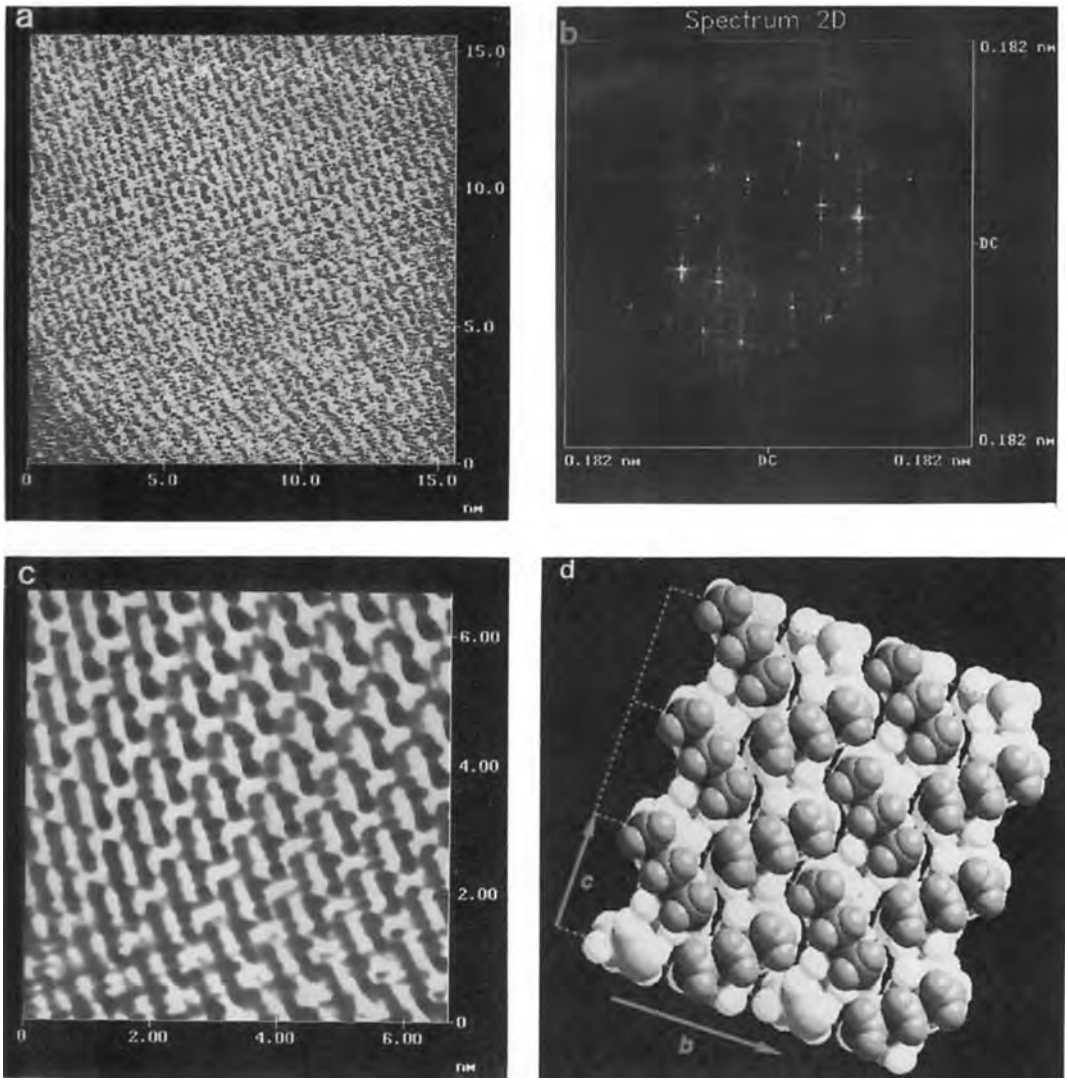


Figure 41. (a) High resolution AFM image of syndiotactic polypropylene, recorded in the liquid cell (force imaging, constant-height mode, no filter, scan rate 57 Hz or $8\,s^{-1}$). Orientation of chain axes is about 1 o'clock. (b) Power spectrum of the upper left part of the picture in (a). Periodicities along and normal to the chain axis direction are 0.72 and 1.1 nm, respectively. Note the appearance of more intensity peaks than in the electron diffraction pattern in Fig. 6, which is linked to the fact that the surface symmetry of sPP differs from its bulk one. (c) Fourier-filtered image of the upper part of (a) produced when selecting 20 pairs of spots in the power spectrum in (b). Note the better resolution of CH_3, CH_2, and CH_3 units oriented nearly normal to the scan direction. (d) Schematic representation of the AFM image in (c) using molecular modeling and shown in correct mutual orientation (cell III in Fig. 1, space group *Ibca*). The atomic van der Waals radii are depicted at 83% of their values to allow better visualization of the individual chains; the uppermost groups visualized in AFM are darkened [175].

investigations of organic molecules have been undertaken in the former situation. For a given tip, the contrast depends on the substrate, the structure and orientation of the absorbed molecules, and the applied bias [178, 179]. Several explanations for the change in measured current have been proposed: (1) changes in the electron density of states within the substrate or tip occur due to their electronic interaction with the absorbate (sample) [180]; and (2) changes in the tunneling barrier or resonant tunneling are caused by electronic levels close to the Fermi level of tip and substrate.

The imaging process can be described in terms of the transfer function

$$I(\boldsymbol{k}) = O(\boldsymbol{k})H(\boldsymbol{k}) + N(\boldsymbol{k})$$

where $I(\boldsymbol{k})$ is the Fourier transform of the image, $O(\boldsymbol{k})$ is the Fourier transform of the object, $H(\boldsymbol{k})$ is the Fourier transform of the transfer function, $\boldsymbol{k}$ is the scattering vector, and $N(\boldsymbol{k})$ is an additional noise term.

In TEM, $H(\boldsymbol{k})$ depends mainly on the spherical and chromatic aberration constants and the defocus value, while $N(\boldsymbol{k})$ depends on the recording device (photographic film or CCD camera). These parameters can be determined and the effect of $H(\boldsymbol{k})$ on the image can be calculated (see Sec. 4.1.2 of this Chapter). The corresponding parameters are not so well known for STM [180], but are certainly related to the geometry of the tip. For conductive surfaces with very small corrugation, the STM transfer function is given by [181]

$$H(\boldsymbol{k}) = \exp\left(\frac{-k^2}{k_0^2}\right)$$

where $k_0 = 4k/(r+s)$, r is the tip radius, s is the spacing between the tip and the sample, and k is the inverse decay length for tunneling [76].

For complex tip and sample geometries the imaging properties become highly nonlinear and harmonics with high amplitudes are created by the imaging process itself [178]. For this reason, analysis of the spatial frequencies in Fourier space can be very misleading. In addition to these geometrical considerations, the forces between the tip and the sample may deform the latter; these forces are approximately 10^{-10} N for organic samples, which is much lower than those for graphite, silver, or gold. With improved understanding of the contrast mechanisms, new insight into interfacial phenomena are to be expected.

In the following selected examples showing STM images of organic molecules, three different situations are distinguished: ultrathin adsorbate layers, organic conductors, and macromolecules.

Ultrathin Adsorbate Layers

For these studies on chemisorbed molecules under ultrahigh vacuum (UHV), small molecules were attached to reactive and atomically flat surfaces in UHV by chemisorption, thus immobilizing them [182]. In this way high resolution images were obtained from arrays of benzene coadsorbed with carbon monoxide on Rh(III) [183] (Fig. 42).

In experiments on *physisorbed molecules at interfaces*, the internal interface between two condensed media is imaged, one being a conducting solid and the other a gas, liquid, or soft solid. In such investigations the basal plane of highly oriented pyrolitic graphite [184] HOPG and MoS_2

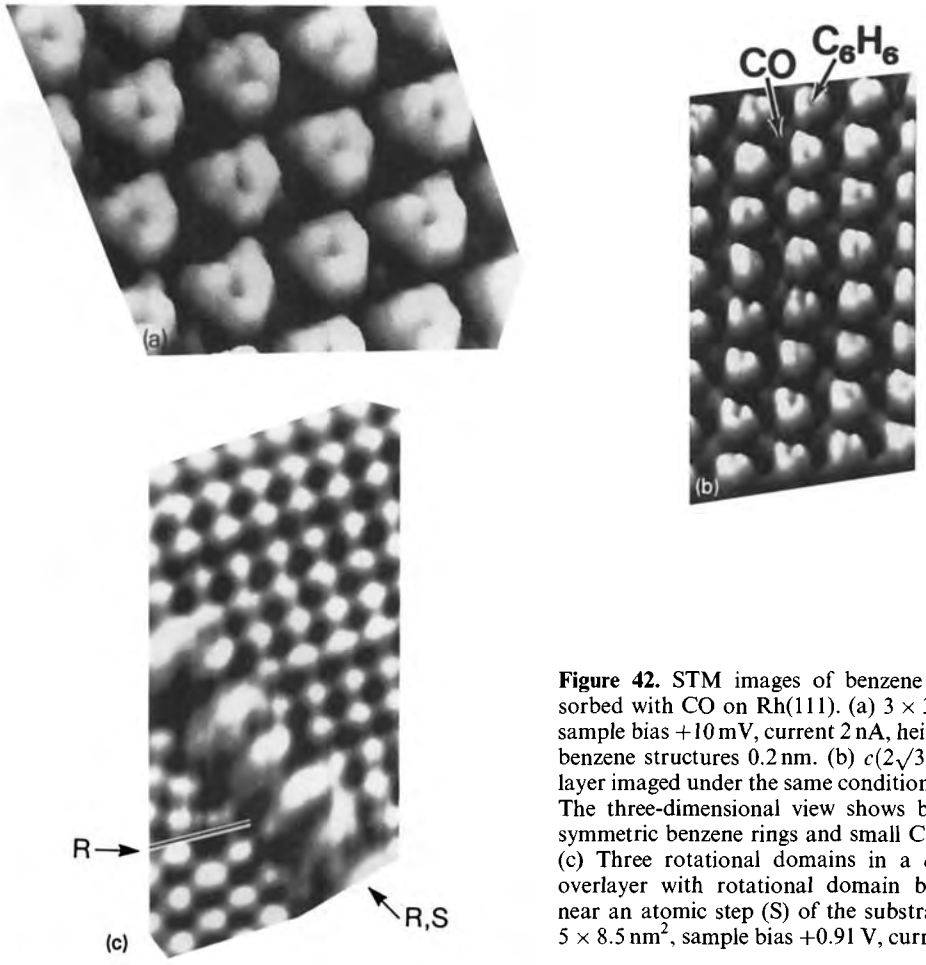


Figure 42. STM images of benzene (C_6H_6) coadsorbed with CO on Rh(111). (a) 3×3 overlayer [8]: sample bias +10 mV, current 2 nA, height of ring-like benzene structures 0.2 nm. (b) $c(2\sqrt{3} \times 4)$rect overlayer imaged under the same conditions as in (a) [17]. The three-dimensional view shows both three-fold symmetric benzene rings and small CO protrusions. (c) Three rotational domains in a $c(2\sqrt{3} \times 4)$rect overlayer with rotational domain boundaries (R) near an atomic step (S) of the substrate: image size $5 \times 8.5\,nm^2$, sample bias +0.91 V, current 2 nA [183].

MOPG [185] was used as substrate. Examples are shown in Fig. 43 [177].

High resolution molecular images at solid/fluid interfaces were thus obtained for two liquid crystals 4-n-octyl-4′-cyanobiphenyl and 4-(*trans*-4n-pentylcyclohexyl)benzonitrile on HOPG [186] and other liquid crystal/HOPG interfaces [187], as well as a homologous series of cyanobiphenyls [188]. The investigations showed that the molecular order at the interface is better than that in the bulk, and that the sample structure depends on the substrate (Fig. 44).

It has also been shown by STM that long-chain alkanes form highly ordered monolayers at the interface between organic solutions on HOPG [189–192].

Organic Conductors

Organic conductors which have been imaged by STM are tetrathiafulvalenetetracyanoquinodimethane [193], β-(BEDT-TTF)I_3 which becomes superconducting at low temperatures [194], organic charge-transfer complexes [195], and conducting polymers such as polyacetylene.

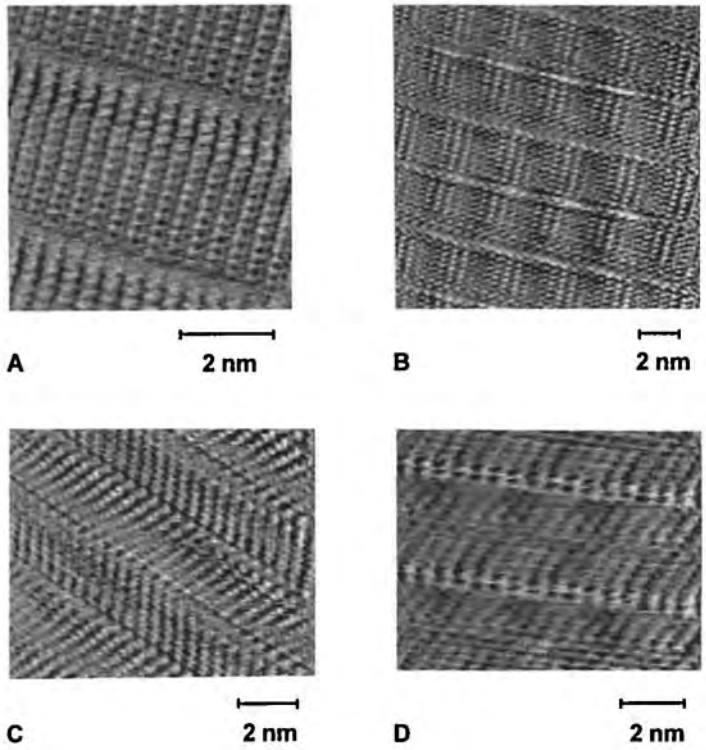


Figure 43. STM images of molecular monolayers adsorbed to the basal plane of graphite: (a) The alkane heptacosane ($C_{27}H_{56}$); (b) the fatty acid stearic acid ($C_{17}H_{35}COOH$); (c) the alcohol octadecanol ($C_{18}H_{37}OH$); (d) didodecylbenzene [$H_{25}C_{12}(C_6H_4)C_{12}H_{25}$]. The brightness is proportional to the current at quasi-constant height. The four cases show alkyl chains perpendicular to ((a) and (b)) and tilted ((c) and (d)) relative to the lamella boundaries (first and second row), as well as chain packings commensurate ((a) and (c)) and noncommensurate ((b) and (d)) with the substrate lattice (first and second column) [177].

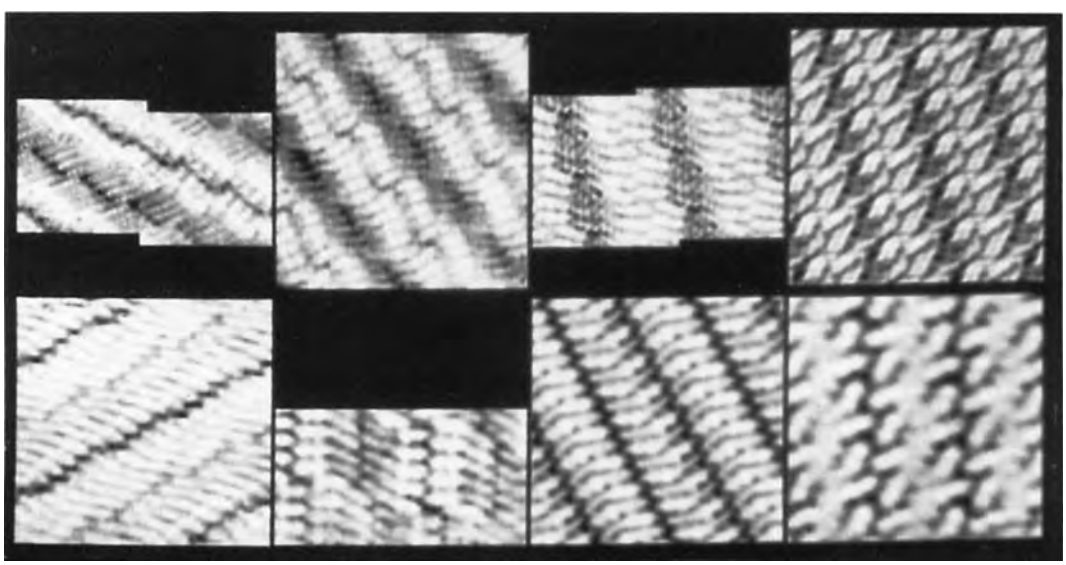


Figure 44. STM images of four alkylated cyanobiphenyls (6CB through 12CB) with 6, 8, 10, and 12 carbon atoms in the alkyl side-chain. The top row displays the images on MoS_2 and the bottom row images on graphite [188].

Macromolecules

Synthetic macromolecules produce STM images which show, at best, molecular resolution. The reason for this is the fact that they are less well ordered than their monomeric analogs and resolution may be adversely affected by chain ends, side groups, etc. The situation is most favorable for stiff macromolecules [196] and polyglutamate [197]. In the case of large biopolymers such as DNA, which are stabilized by intermolecular forces, good images can be obtained which clearly show the right-handed helical structure [198]. Another class of biopolymers giving extremely good STM images are membrane proteins [199].

13.3.3 Electron Spectroscopic Imaging and Electron Energy-Loss Spectroscopy

The simultaneous imaging of the crystalline and amorphous areas of semicrystalline polymers is possible without a staining agent by using electron spectroscopic imaging (ESI) in an energy filtering electron microscope (EFTEM). The method has been applied successfully to obtain structural information about polyethylene [200, 201], block copolymers such as poly(styrene-b-methylphenylsiloxane) [202] and polystyrene-block-poly(2-vinylpyridene) [203], using the Zeiss EM 902 at 80 kV for imaging and the Zeiss EM 912 W operated at 120 kV to obtain electron energy loss spectra. The inelastic interactions of the primary beam with inner shell electrons of the sample give rise to distinct energy losses, the magnitudes of which are characteristic for each element. Conventional electron energy-loss spectrometers contain an electromagnetic prism, a slit diaphragm for energy selection and an electron detector to measure intensity. The local resolution is limited to about 500 Å. Calibration is difficult because of the remanence of magnetic fields. In contrast, in the electron energy-loss spectrometer of the prism–mirror–prism type [204] the spectrum is scanned by varying the accelerating voltage and the electron optical system remains fixed, so that fine calibration of the magnetic prism is unnecessary. By use of a slit diaphragm in the energy dispersive plane, 'monochromatic' electrons can be selected for image formation. The potential advantages of ESI were demonstrated in an investigation of the morphology of polyethylene fibers crystallized from stirred xylene solutions. While it was clear from early conventional TEM studies that the resulting 'shish-kebab' structures consisted of a central core with chain-folded lamellae attached to the central backbone, details relating to the microstructure of the core still remained to be elucidated (Fig. 45). Using stepwise solvent exchange followed by low-temperature embedding in Lowicryl HM20, the polyethylene crystalline gel was changed to the solid state without morphological deformation and excellent high contrast images obtained (Fig. 46) [203]. The contrast arises from the higher density of carbon atoms in the crystalline areas compared with the Lowicryl matrix.

Electron energy-loss spectroscopy (EELS) is particularly well adapted for investigating block copolymers if staining is difficult or undesirable. A good example which shows the improvement in contrast

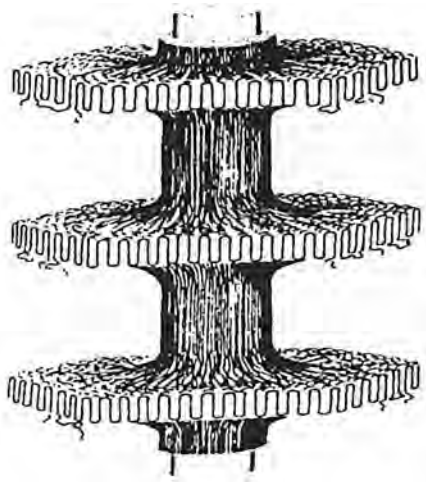


Figure 45. Schematic 'shish-kebab' structure of polymer fibers.

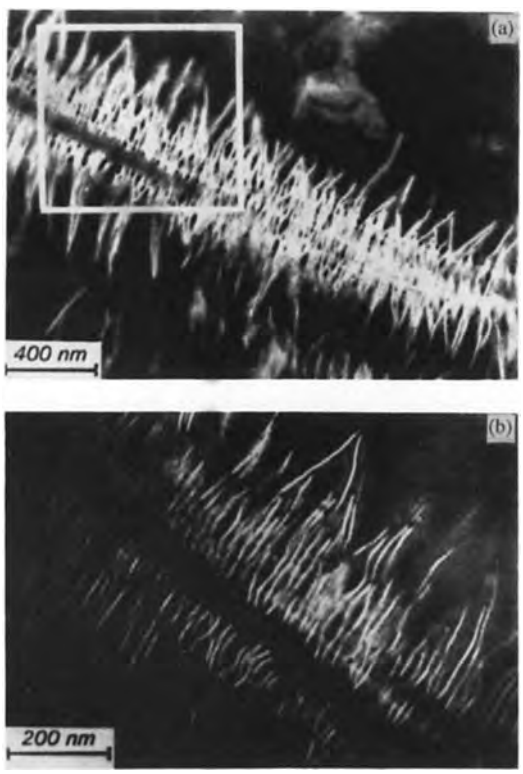


Figure 46. (a) High-contrast ESI image of nonisothermally crystallized polyethylene shish-kebab; (b) detailed view of the marked area from (a) showing the polyethylene backbone.

is a polystyrene-block-poly(1-methyl-2-vinylpyridium iodine) copolymer, and the iodine-specific image with $\Delta E = 206\,eV$ (Fig. 47). This iodine-specific energy loss peak was found to be particularly useful as there was only slight overlap with other peaks, while other and stronger iodine-specific loss peaks were found to be adjacent to nitrogen or carbon loss peaks [205].

13.3.4 Low Voltage High Resolution Scanning Electron Microscopy

Most polymers are nonconducting, so that in conventional SEM operation charging effects necessitate coating samples with a 200–300 Å metal film, thus obscuring fine structural details. At conventional accelerating voltages (≈30 keV) the beam/sample interaction volume in carbonaceous materials is about 10 μm in diameter. Under these conditions, secondary (inelastically scattered) electrons emerge from a shallower, more localized region than the (elastically) backscattered electrons, and therefore provide better image resolution. The secondary image resolution so obtained is effectively equal to the diameter of the incident electron probe, which in turn depends on the brightness of the electron source and the accelerating conditions. Thus the highest resolution which can be obtained for organic samples with a conventional scanning electron microscope (LaB_6 tip at ~30 keV) has been about 50 Å. With a field emission gun and a low voltage (~1–2 keV), the beam can be focused to about 25 Å and even to <10 Å at higher energies (>10 keV).

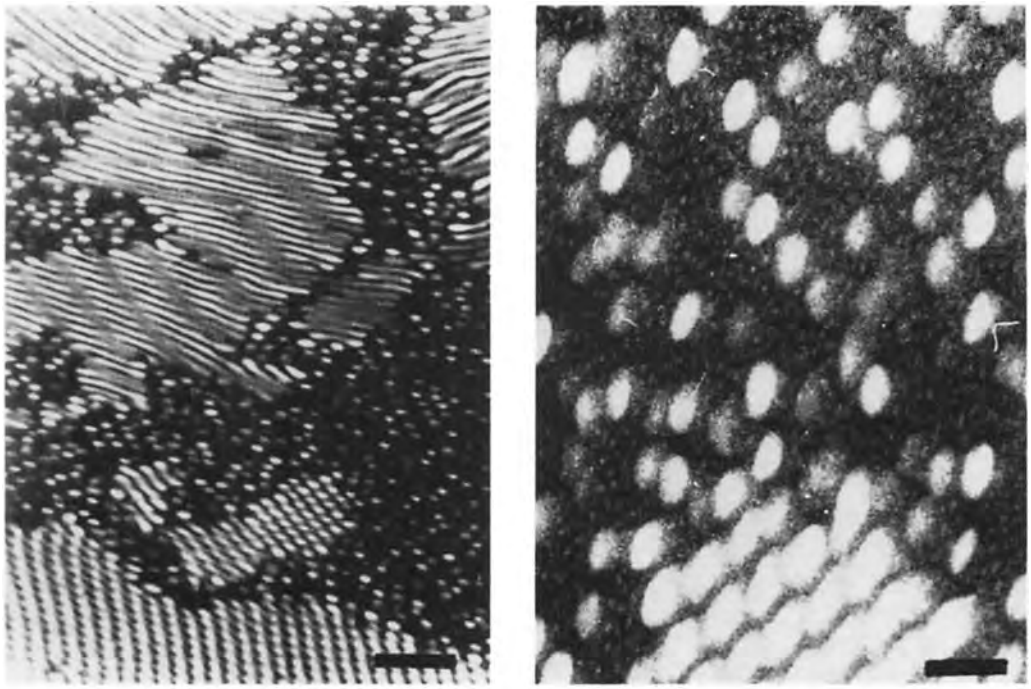


Figure 47. Ultrathin section of a polystyrene-block-poly(1-methyl-2-vinylpyridinium iodide) film from *N*,*N*-dimethylformamide at 50°C. Iodine-specific image (energy loss compensation $\Delta E = 206\,\text{eV}$); scale bar (a) 400 nm, (b) 100 nm [205].

The situation for carbonaceous materials has been recently reviewed [206]. The secondary electrons escape from a depth of about 200 Å, regardless of incident beam energy. Therefore, the only way to improve resolution is to use an accelerating voltage so low that the interaction teardrop dimensions are less than the 200 Å escape depth. The resolution is decreased, but charging is eliminated. When the secondary electrons escape depth is of the same order as the incident electron penetration depth, the build-up of surface electron charge is eliminated. At a given accelerating beam energy, usually called $E2$, the sum of the secondary and backscattered electrons equals that of the incident beam and it is possible to obtain images without a conductive surface coating. The value of $E2$ was found to be in the range 0.9–1.5 keV for different polymers [207, 208]. Because polymers have different $E2$ values, secondary electron imaging compositional contrast can be obtained from copolymers, which also give rise to surface topography.

In practice, charging is not completely eliminated even when working at the crossover voltage, because local variations always occur [206]. Another practical difficulty arises because often samples can be imaged uncoated at TV rates on the cathode ray tube (CRT), but an image cannot be collected because the slower scan rates required to record an image increase charging. There are two possible solutions:

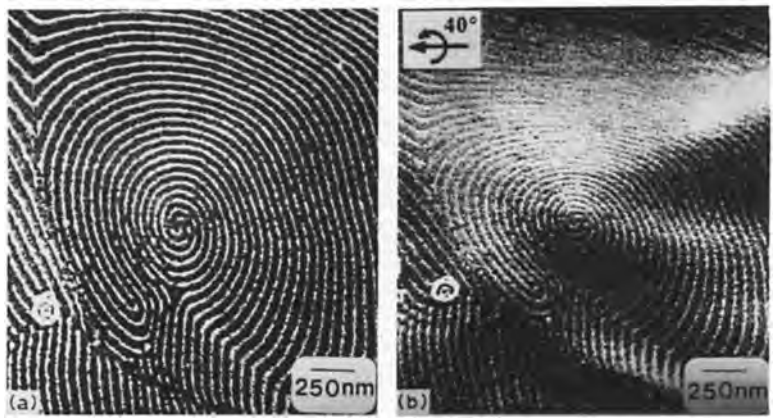


Figure 48. Low-voltage, high-resolution SEM images of an OsO_4-stained PS-PB lamellar diblock copolymer: (a) incident beam approximately parallel to the regional surface normal, giving maximum contrast; (b) the upper region of the micrograph shows a loss of contrast and change in apparent lamellar spacing due to the 40° sample tilt. (From Thomas [205].)

(1) a frame-grabber can be used to take the digital output of the CRT and integrate several images taken at fast scan speeds; and (2) new coating technologies, such as dual ion beam sputter coaters, enable extremely thin continuous conductive coatings of about 10 Å to be applied [209]. Many technical aspects regarding the reduction of contamination and improvement of resolution as well as contrast mechanisms are discussed in the article by Thomas [205], from which Fig. 48, showing a polystyrene–polybutadiene diblock copolymer, is reproduced.

13.4 Electron Crystallography

Electron crystallography is known to pose so many difficulties that its success in solving structures in the most difficult samples of all, namely beam sensitive organic materials, seems very surprising. Undoubtedly, benefit has been obtained from the enormous development in X-ray methods [210], but the technique is still far from reaching the resolution and almost routine application common in X-ray structure analysis and refinement. Unfortunately, many long-chain polymers form only microscopically small crystals, which are quite unsuitable for X-ray analysis, and so electron diffraction is the only possible route to structure determination. The technique has the enormous advantage that selected area diffraction produces single-crystal diffraction patterns which can be obtained in different projections with the aid of a goniometer stage or by using epitaxial orientation. A list of the diffraction structure analyses made of linear chains up to 1993 is given in Table 1 [211–242]. In the older papers, either only the unit cell constants were determined, or a crystal structure which had already been determined by X-ray analysis was confirmed. However, both statistical methods and computational facilities have improved considerably, so that ab initio structure determinations from electron diffraction data are now possible. These methods are described in more detail below. After some experimental information, the discussion is divided into three additional subsections:

- Experimental procedure.
- Simulation of electron diffraction patterns and high resolution images.

Table 1. Quantitative electron diffraction structure analyses of linear polymers (from Dorset [211])

Polymer	Data	*R*	Phasing technique
Two-dimensional determinations			
Poly(tetrafluoroethylene) (oligomer) [211]	*hk*0	0.11	Theoretical model
Poly(diacetylene) [212]	0*kl*	0.12	Patterson map and chain rotation
trans-Poly(acetylene) [213]	0*kl*	0.09	Chain rotation
Poly(ethylene sulfide) [214, 215, 216]	*hk*0	0.33 (0.19 when corrected)	Conformational refinement
α-Poly(3,3-bis-chloromethyl oxacyclobutane) [217]	*hk*0	0.25	Conformational refinement, previous model
Cellulose triacetate [218]	*hk*0	0.26	Conformational refinement
Poly(trimethylene terephthalate) [219]	*hk*0	0.35 (0.25 when corrected)	Conformational refinement
Poly(hexamethylene terephthalate) [220]	*hk*0	0.17	Conformational refinement
Poly(1,11-dodecadiyne) macromonomer	*h*0*l*	0.13	Conformational refinement
Cross-linked poly(γ-methyl-L-glutamate) [221]:	*h*0*l*	0.13	
α-form [222]	*hk*0	Not given	Theoretical model
β-form [223]	*h*0*l*	0.38	Theoretical model
Dextran:			
Low temperature [224]	*hk*0	0.26	Conformational refinement
High temperature [225]	*hk*0	0.18	Conformational refinement
Anhydrous nigeran [226]	*hk*0	0.25	Conformational refinement
Chitosan [227]		<0.20	Conformational refinement
β-Isotactic polypropylene [228]	*hk*0	0.10	Model building
Three-dimensional determinations			
Polyethylene [229, 230]	*hkl*	0.21	Rotational search, direct methods
Poly(ϵ-caprolactone) [231, 232]	*hkl*	0.20	Fiber X-ray models, direct methods
Poly(sulfur nitride) [233]	0*kl* + *hk*0	0.19	Patterson function
cis(Polyacetylene) [234]	*hkl*	0.13	Rotational refinement
Valonia cellulose [235]	*hkl*	0.23	Conformational refinement, direct methods
Mannan I [236, 237]	*hkl*	0.22	Conformational refinement, direct methods
V_H amylose [238]	*hkl*	0.24	Conformational refinement
Poly(*trans*-cyclohexanediyldimethylene succinate) [239]	*hkl*	0.24	Conformational refinement, direct methods
γ-Poly(pivalolactone) [240, 241]	*hkl*	0.14	Conformational refinement
Poly(1-butene), form III [242]	*hkl*	0.26	Tangent formula minimal principle

- Direct methods.
- The maximum entropy method.

13.4.1 Experimental Procedures

13.4.1.1 Sample Preparation

The following methods of sample preparation can be used:

- Growth of single crystals from solution.
- Epitaxial growth on a suitable host crystal [243].
- Application of shear forces, either by spin coating or rubbing.
- Orientation in a liquid crystalline phase by electric or magnetic fields [149].
- Orientation in mono- or multilayers by the Langmuir–Blodgett method [244, 245].

13.4.1.2 Electron Diffraction and High Resolution Imaging

Electron diffraction patterns must be obtained in several projections by a tilting series, using the goniometer stage. Experimentally this step is very demanding. The beam dose has to be reduced to a minimum and, at the same time, the tilting axis carefully aligned in order to obtain diffraction patterns from suitable zones. In some aromatic samples, as many as nine different projections can be obtained [246].

From these experiments the unit cell and space group are determined. Alternatively, a suitable preparation technique may be used to obtain the desired orientation. In order to check whether forbidden reflections are occurring, it is desirable to compare data with a simple X-ray powder pattern. Using the series of electron diffraction patterns (and, if possible, the X-ray powder pattern) the basic zones must be established and consistent indexing achieved in all zones [246, 247].

High resolution images are obtained by phase contrast imaging using cryogenic and low-dose facilities. It is convenient if the phase contrast transfer function can be controlled by transferring the image via a video camera to a computer with fast Fourier transform capabilities [159, 167].

13.4.1.3 Quantitative Analysis of Electron Diffraction Patterns and High Resolution Images

The quantitative analysis of electron diffraction patterns represents a crucial and difficult step. At this stage the space group is known and undesirable symmetry-forbidden reflections are deleted. At present the most convenient analysis of the intensities is performed by using a CCD camera and transferring the data to a PC where it can be analyzed using the ELD software package [248]. Undoubtedly, if CCD cameras with a large dynamic range could be used for on-line analysis of diffraction patterns, this would be of great benefit in electron crystallography. A possible method of analyzing the images and confirming the plane groups in the appropriate projection is available in the CRISP software [249]. Structure factor magnitudes from the electron diffraction patterns are determined in the usual way. The high resolution images are digitized by a CCD camera and transferred to a PC via a frame-grabber. The diffraction pattern of the image is calculated by a fast Fourier

transform routine and the amplitudes and phases determined as in older programs developed at the MRC Laboratory in Cambridge [250]. From this information CRISP finds and imposes the correct symmetry of the crystal and the image is thus restored. These images can then be compared with those obtained from the model structures, as described below.

13.4.2 Simulation of Electron Diffraction Patterns and High Resolution Images

13.4.2.1 Molecular Conformation

To begin simulation, the molecular conformation has to be calculated. This is now standard practice in theoretical chemistry and many degrees of accuracy can be attained, ranging from force field calculations via semiempirical quantum mechanical methods to the ab initio methods. In view of the fact that the molecular conformation as calculated for the gas phase is not identical to the one adopted in the crystal field, the most accurate ab initio methods need not be used. Instead, most of the calculations can be performed using the semiempirical methods incorporated in MOPAC 6.0. The approximations used in semiempirical quantum mechanical calculations are well known [252–254]. All semiempirical quantum mechanical methods use the formalism of ab initio methods (linear combination of atomic orbitals (LCAO) self-consistent field (SCF) approach). The methods differ only with respect to the type of approximations used (e.g. complete neglect of differential overlap (CNDO), modified intermediate neglect of differential overlap (MINDO), modified neglect of differential overlap (MNDO), extended Hückel approach). Other models differ only in the parametrization (MNDO, Austin model (AMI), parametric method 3 (PM3)). The method has been used to obtain the structure of monomeric analogs of the discotic triphenylene ethers and esters [159, 160] as well as the sanidic aromatic polymeric amides [164, 165] discussed in Sec. 13.2.5.

Most molecules have several local energy minima. The conformation which is closest to the one which gives good agreement between simulated and experimental diffraction pattern is chosen to initiate the analysis. See, for example, the cyanobianthryl molecule, which was analyzed in order to determine its suitability for non-linear optical applications [246, 251] (Fig. 49).

13.4.2.2 Simulation of Diffraction Patterns

The unit cell and possible space groups are calculated from a series of experimental diffraction patterns obtained in different projections. Generally, several space groups can be postulated initially, on the basis of the observed extinctions. This number can be reduced when the symmetry requirements of the molecule are taken into account. For simulation, the MOPAC calculated molecule can be placed into the unit cell using CERIUS such that the required symmetries of the molecule and the cell are satisfied and agree with the observed extinctions and symmetry. As the molecule is adjusted, the changing diffraction pattern is immediately displayed; a first qualitative agreement and a good value of the density can usually be

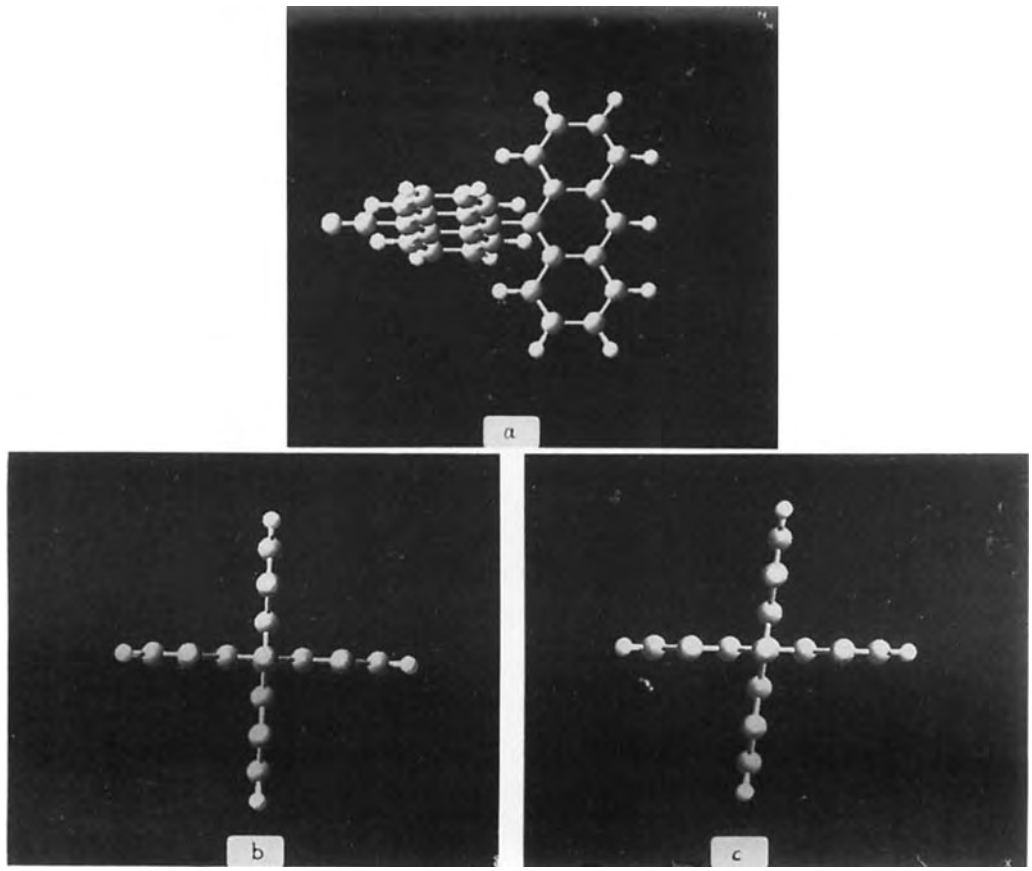


Figure 49. Conformation of CNBA molecule: (a) showing the angle between anthracene moieties; (b) in the gas phase; (c) in the crystalline phase [246, 251].

achieved within reasonable time. At this stage, the packing energy is usually positive and a number of nonallowed close contacts observed. These require adjustment of distances in the Z matrix to satisfy specific criteria involving the packing of aliphatic chains or known bond angles from molecular fragments. Subsequently, the CRYSTAL PACKER is used to ensure that the packing energy of the crystal is negative and that nonallowed contacts are eliminated while retaining the correct symmetry. The potential energy of the crystal is written as the superposition of various two-body, three-body, and four-body interactions:

$$E = E_{\mathrm{VdW}} + E_{\mathrm{Coul}} + E_{\mathrm{hb}} + E_{\mathrm{tors}}$$

The van der Waals term E_{VdW} is treated using the Lennard–Jones functional term and will, of course, represent the major contribution in organic crystals. The Ewald summation technique is used to treat the Coulomb energy E_{Coul} [255]. The energy of the hydrogen bonds E_{hb} is treated with the CHARM-like potential.

A Dreiding force-field is used for the calculation of subrotation interactions E_{tors} [256]. Finally, the R factor, which relates the observed and calculated intensities by the relationship

$$R = \frac{\sum_{hkl} ||F_0| - |F_c||}{\sum_{hkl} |F_0|}$$

has to be calculated. For electron diffraction, the values commonly achieved in X-ray diffraction cannot be reached; a reasonable value, if all the above criteria are satisfied, lies in the range 0.1–0.3. Examples of successful simulations of organic monomeric analog crystals based on these considerations are the dione head group [247], the triphenylene ether [159], and the cyanobianthryl [246] (Fig. 50). Further examples are polybutene (Fig. 51) [257], several polysaccharides [258], and chitosan [259].

It is pertinent to ask the question why simulation methods are proving to be successful with organic molecules, despite the difficulties associated with beam damage, as well as dynamical and secondary scattering effects. It seems that there are a number of contributing factors. The unit cell is not too large by biological standards ($\sim$1000–2000 Å^3), containing only a few hundred atoms. Of these, many are covalently linked, with well-defined bond lengths. Furthermore, parts of the molecule frequently have very stiff, well-defined fragments. In addition, aliphatic chains often adopt a close-packed arrangement similar to that in alkanes, so that the attached functional groups are forced to fit into available spaces by molecular subrotations, while keeping allowed intra- and intermolecular bond distances and retaining the symmetry observed in the electron diffraction patterns.

13.4.2.3 Simulation of Images

The images can be simulated using the atomic coordinates from the model structures obtained by electron diffraction simulations. The CERIUS HRTEM mode simulates dynamical diffraction patterns and real-space images using the multislice method [260]. The expression for the two-dimensional image intensity distribution $I(x, y)$ is given by:

$$I(x,y) = |q(x,y)|^2 \cdot \left[\frac{J_1(\pi u_0 r)}{\pi\tau}\right]^2 \times |F \exp\{i\chi\}|^2$$

where $[J_1(\pi u_0 r)]/\pi r$ is the first-order Bessel function representing scattering from a spherical atom with coordinates x, y at a distance r in real space, q is the transmission function and $|F \exp\{i\chi\}|$ is the Fourier transform of the transfer function χ where

$$\chi_{ph} = \frac{2\pi}{\lambda}\left(\Delta f \frac{\boldsymbol{k}^2}{2} - C_s \frac{\boldsymbol{k}^4}{4}\right)$$

$\boldsymbol{k}$ is the scattering vector, C_s is the spherical aberration, Δf is the defocus, and λ is the electron wavelength.

In practice, for organic samples the required transfer function χ is carefully determined beforehand and adjusted in an adjacent area before the area of interest is moved into the beam and immediately photographed.

Examples of successful structure determinations using a combination of electron diffraction, high resolution imaging, and molecular modeling are the liquid

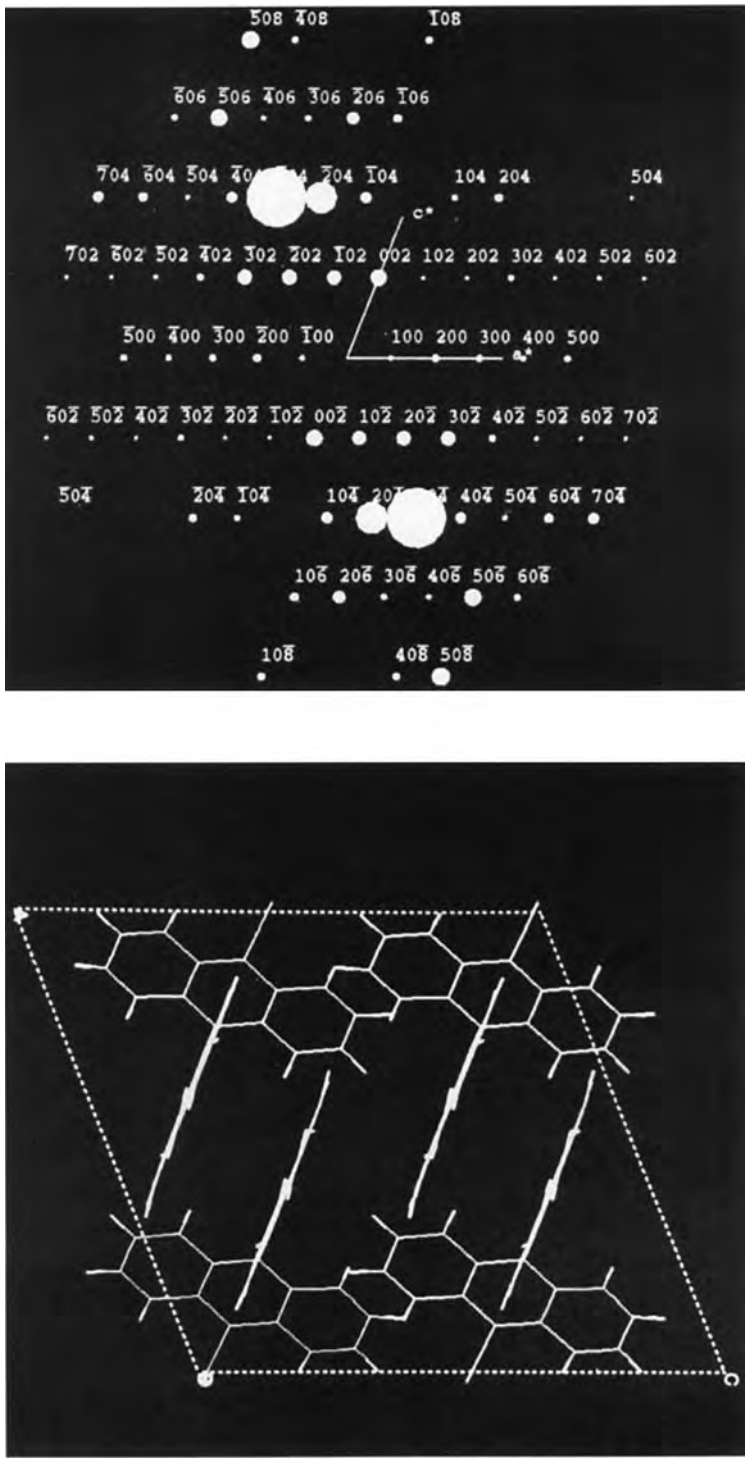


Figure 50. Model and its diffraction pattern of cyanobianthryl obtained by simulation of experimental diffraction patterns [251].

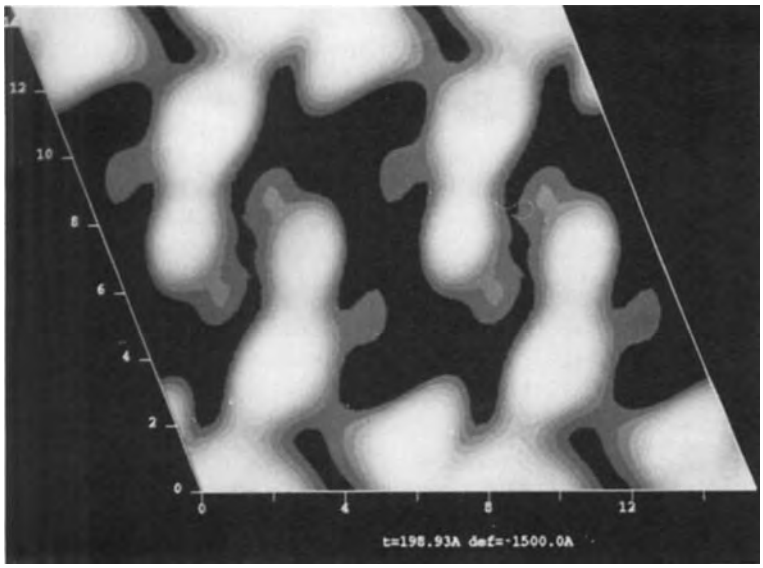


Figure 50 Continued

crystalline polymers and their monomeric analogs [159, 160, 164, 167, 168], as well as several linear polysaccharides [224, 225, 236, 244].

13.4.3 Direct Methods

For several years direct methods have been used successfully for the analysis of organic crystals from electron diffraction data [261–267]. In the case of large macromolecules, electron diffraction structure factor magnitudes have been used together with crystallographic phases determined from high resolution images in order to visualize the unit cell contents [268] using appropriate methods of averaging statistically noisy images [269] to improve their interpretability. However, in the case of organic polymer crystals with small molecules and small unit cells the direct phasing methods can be used for ab initio structure determination at very high resolution because electron diffraction patterns can be obtained with a resolution down to about 1 Å. Initially, normalized structural factor magnitudes $|E_h|$ are calculated using electron scattering factors F_h

$$E_h^2 = \frac{|F_h|^2}{\epsilon \sum_{i=1}^{N} f_i^2}$$

where f_i is the atomic scattering factor for atom i, and ϵ is a multiplicity factor which accounts for translational symmetry elements.

In direct methods the probability is calculated that certain combinations of crystallographic phases ϕ_h will have a specific value ψ (0 or π in centrosymmetric crystals) [270, 271]

$$\psi = \Phi_{h_1} + \Phi_{h_2} + \Phi_{h_3} + \cdots$$

subject to the constraint

$$\sum_i h_i \equiv 0$$

where h_i are the Miller indices of several reflections, based on normalized structure

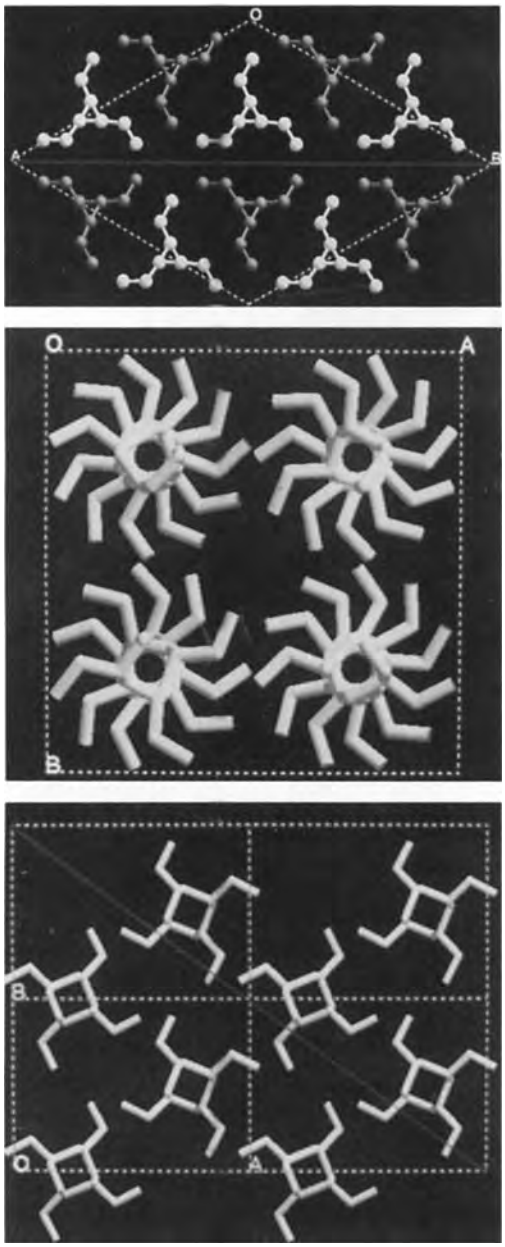


Figure 51. (Top) computer-generated molecular model of PBu: form I and I′ unit cell; (*ab*) plane; View along chain axis showing 3_1 helices in trigonal packing. Note the layered structure with interdigitated antichiral helices. (Middle and bottom) computer-generated molecular model of PBu: form III unit cell; (100) projection showing the 4_1 helices in projection (orthorhombic packing) [257].

factor magnitudes. Initially, a subset of highly probable phases is chosen and combined with a small number of additional individual phases to give a consistent set of simultaneous equations. The process is repeated with different combinations of reflections. Often a consistent set of phased structure factors can be found and some features of the molecular packing in the unit cell determined from the potential map obtained by Fourier transformation [272]. The intensity values can be improved when the effects of dynamical scattering and crystal bend deformation are considered [273, 274].

In principle, the phases obtained from the Fourier transform of high resolution images can be used to obtain some of the phases from the low resolution diffraction maxima for the initial subset [275], but they may be misleading if the symmetry of the object is not retained fully during the imaging process, so that it is better to attempt an ab initio determination.

Once an initial crystal structure has been found, the valence parameters (bond lengths and angles) can be optimized using various different techniques.

Difference Synthesis

Given the potential map derived from the phased structure factors ϕ_c, a difference synthesis,

$$\phi_0 - \phi_c = \frac{1}{V}\sum\sum\sum \Delta F_{hkl}(-2\pi i r \cdot s)$$

can be calculated, where $\Delta F = (F_0 - F_c)$, the difference between observed and calculated structure factors [276].

Least-Squares Fit of Parameters

In this approach the atomic positions, thermal parameters, etc., are adjusted to

give best agreement between the calculated and experimental data. In practice, the multisolution phase determining programs MULTAN and RANTAN have proved effective for phase extension [277], on the condition that the subset phases are known from other experiments. The program QTAN was applied to the electron diffraction problem with more success [279, 280]. In this procedure the most likely phase solution is sought in the most negative value of NQEST [279] based on negative quartet invariants. The results of direct phase determination for the polymers polyethylene, poly(e-caprolactone), and poly(1-butene) are shown in Fig. 52 [277]. The projection of polybutene (Fig. 53) shows excellent agreement with the structure obtained by simulation (Fig. 51).

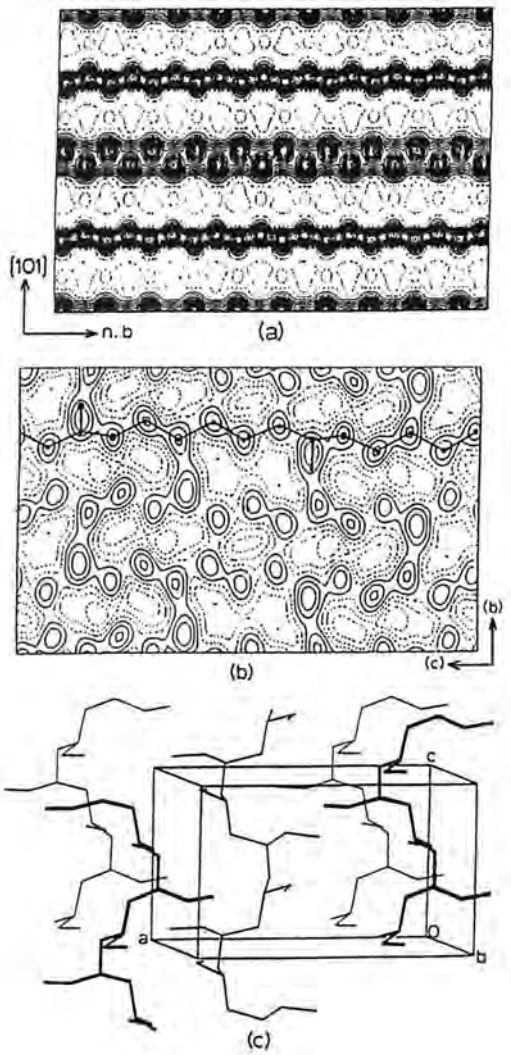


Figure 52. Results of direct phase determination for three polymers: (a) polyethylene; (b) poly(ε-caprolactone); (c) poly(1-butene) [278].

13.4.4 The Maximum Entropy Phasing Method

Direct methods of crystal structure determination impose rather stringent requirements on the data set:

- A data resolution of 1.1–1.2 Å.
- A nearly complete sampling of reciprocal space to this resolution.
- A reasonably accurate set of intensity measurements devoid of large systematic errors.

The third criterion may be particularly troublesome because of secondary and dynamic scattering. The maximum entropy approach is not so constrained with respect to data quality and has been used successfully to solve the structure of two monomeric analogs containing the chromophores for second-harmonic generation [246, 247] and the structures of some small organic molecules [281, 282]. The procedure is summarized below.

In direct methods, probability limit theorems are used to estimate the joint probability distribution of suitably chosen structure factors; the substitution of the

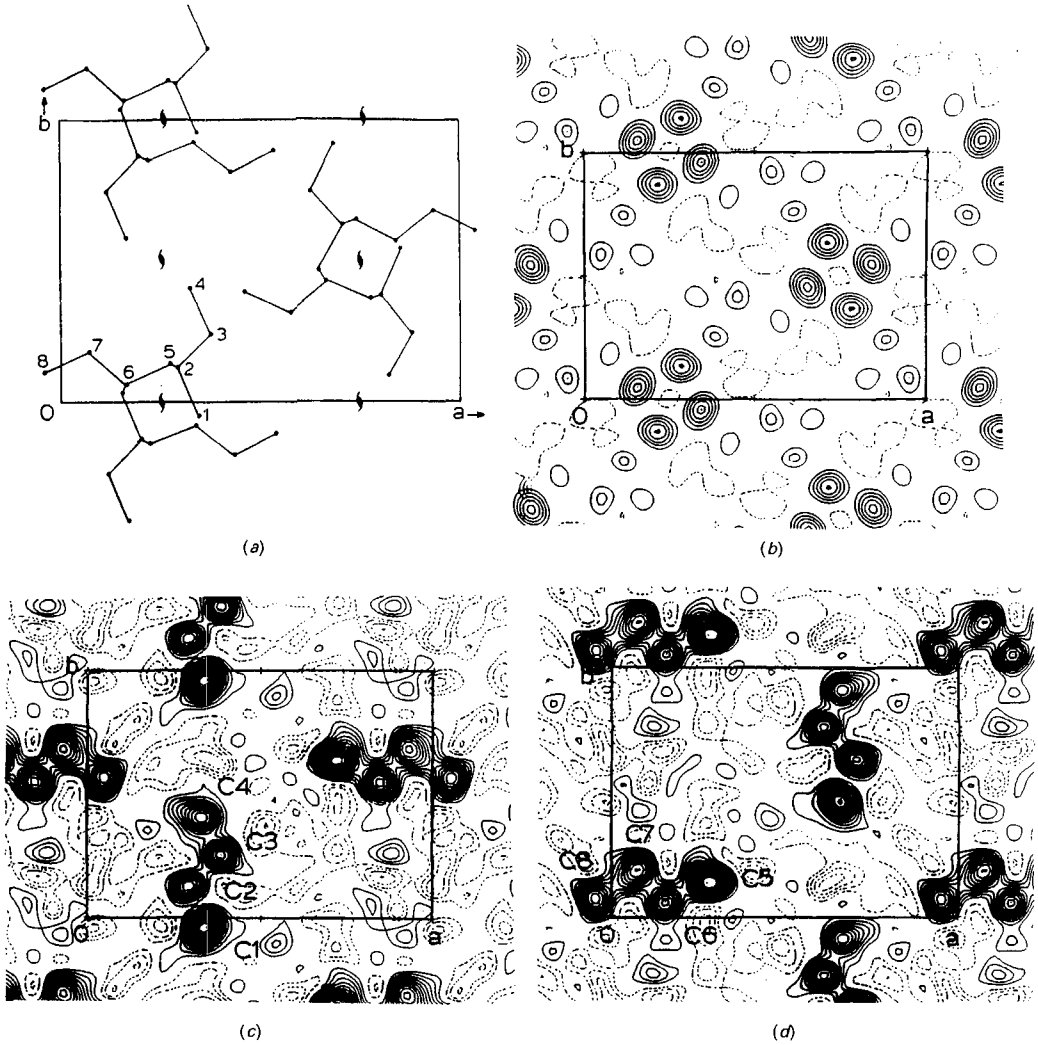


Figure 53. (a) Molecular packing in the projection along c, showing the pseudo-four-fold symmetry. (b) Electrostatic potential in the projection along c. (c) Electrostatic potential in the section $z = 0.12$. (d) Electrostatic potential in the section $z = 0.35$ (indicating that density persists somewhat beyond the $z \pm 0.25$ required by the pseudo-4_1 symmetry operation) [278].

observed structure factor into these distributions yields conditional joint distributions for the phases, indicating that certain combinations of phase values are more probable than others once the amplitudes are known. However, the traditional procedure involving the Gram–Charlier or Edgeworth series [283, 284], or related expressions [285], is not used in the maximum entropy approach. Instead, the saddlepoint method [286] is invoked. This is equivalent to requiring the distribution of random atomic positions to be updated whenever phase assumptions are made, so as to retain maximum entropy under the constraints embodied in these assumptions. This approach to the phase problem leads naturally to a general multisolution strategy of structure determination [287–289] in which the space of hypothetical

phase sets is explored in a hierarchical fashion by building a phasing tree. Each trial phase set is represented as a node on the tree, and is ranked according to the log-likelihood gain (LLG), or the Bayesian score, associated with it. This, in turn, measures how well a trial phase choice predicts the pattern of observed, unphased intensities via the process of maximum entropy extrapolation. The growth of the phasing tree is guided and controlled throughout by this heuristic function.

For the purposes of ab initio phasing from electron diffraction, the method is implemented as follows [289, 290]:

(1) The unique intensities are normalized using MITHRILL [291, 292] to give unitary structure factors $|U_h^{obs}|$ and their standard deviations σ_h by using electron scattering factors and imposing a suitable isotropic temperature factor.

(2) An origin is defined by fixing phases of suitable reflections using the MICE software with a second neighbor search procedure [293, 294]. These phased reflections form a basis set $\{\boldsymbol{h}\}$. The remaining reflections comprise a disjoint set $\{\boldsymbol{k}\}$ $(\boldsymbol{h} \cap \boldsymbol{k} = \phi)$.

(3) These phased reflections are used in an entropy-maximization procedure using an exponential modeling scheme [289] to procedure an initial maximum entropy atom distribution $q(x, y)$. In this way node 1 of a phasing tree is generated.

(4) At this stage, maximum entropy extrapolation from the Fourier transform of $x(x, y)$ is weak. Therefore further reflections $(h \in k)$ are added on the basis of (a) strong intensity and (b) optimal enlargement of the basis set. This gives rise to a series of phase choices, each of which is represented as a node on the second level of the phasing tree. Each node is subjected to constrained entropy maximization and an LLG is calculated for each one. A Rice-type likelihood function is used to rank the models [287–289].

(5) The LLGs are analyzed for phase indications and only those consistent with *t*-test results [293] are kept. Further reflections are permuted and a new level of nodes generated. The ME distributions associated with the various nodes are not electron potential maps in the traditional sense; instead, centroid maps are produced [289] using a Sim-type filter [294].

The method has been successful with electron diffraction data sets for perchlorocoronene [295] and a variety of organic and inorganic molecules [296]. Examples of a successful application of the method to two organic molecules with functional groups, which had also been analyzed using the simulation method, are the dione head group [247] and cyanobianthryl [246, 251] (Fig. 54). The *ac* projection shown in Fig. 54 demonstrates the excellent agreement with the results of the simulation shown in Fig. 50.

13.5 High Resolution Transmission Electron Microscopy

High resolution imaging of polymeric materials presents a number of specimen-specific problems. However, in the past 10 years there have been some critical improvements in modern microscopes and computer technology which have led to a dramatic increase in high resolution TEM investigations of polymers.

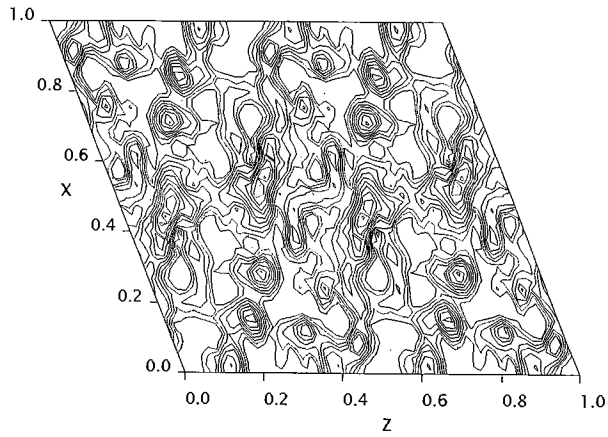


Figure 54. Maximum entropy calculated potential map of 10-cyano-9,9′-bianthryl from electron diffraction data in the [010] projection [246].

The theoretical and experimental conditions required for high resolution imaging are well known [297, 298]. This section is therefore devoted to technical problems concerning organic molecules and their solution (Sec. 5.1 of this Chapter) and the direct impact of such investigations on the understanding of polymer structures (Sec. 5.2 of this Chapter).

13.5.1 Specimen-Specific Problems

13.5.1.1 Radiation Damage

The most serious problem limiting the quality of high resolution images from polymers is beam sensitivity. Radiation damage in polymers during electron-beam exposure is a well-known problem which was first analyzed about 30 years ago [299, 300]. The effect depends on details of the interaction between the electrons and the molecule, which are temperature dependent. In some samples free radicals are formed which move along the polymer chain and finally cause cross-linking (increasing the molecular weight). Alternatively, high energy electrons can cause degradation of the molecule (decreasing the molecular weight). Beam damage is characterized by a quantity called the 'critical end-point dose' J_C and is obtained from the loss of intensity I of a Bragg reflection as a function of dose:

$$I = I_d \exp\left(\frac{-J}{J_C}\right) + I_b$$

A related quantity is J_e, the incident electron dose required to destroy crystallinity in the sample. Typical values at room temperature are 600 electrons nm^{-2} for an aliphatic polymer such as polyethylene and 144 000 electrons nm^{-2} for an aromatic polymer such as poly(paraphenylene benzobisthiadiazole). Therefore the experimental situation is considerably more favorable in the latter case.

Experimentally, it is often necessary to use low-dose and cryogenic methods. These have the disadvantage that a sample position is imaged which is adjacent to the one used for focusing and may therefore require different imaging conditions. In addition, the resolution attainable in good cryogenic holders is only about 5–10 Å.

13.5.1.2 Choice of Microscope Transfer Function

In polymeric samples the features that one wants to image may lie between 2 and 30 Å. Because of beam sensitivity, the correct transfer function cannot be found experimentally by defocusing but has to be calculated beforehand [151, 154, 167]. It is well known that the phase contrast function is given by $\sin \chi$, where

$$\chi(\boldsymbol{k}) = \frac{2\pi}{\lambda}\left[-\frac{1}{4}C_s\boldsymbol{k}^4 + \frac{1}{2}\Delta f\boldsymbol{k}^2\right]$$

λ is the electron wavelength, C_s is the spherical aberration, $\boldsymbol{k}$ is the scattering vector, and Δf is the defocus value.

This function oscillates strongly when the contrast function is manipulated to give a maximum in the small angle region, transferring higher spatial frequencies with positive and negative contrast. Since the image intensity $I(x_1, x_2)$ is given by a convolution $(*)$ of the Fourier transform of the transmission function with the Fourier transform of the transfer function, it is essential to know the transfer function for image interpretation, since in the weak phase approximation:

$$I(x_1, x_2) = 1 + 2\sigma\Phi\varphi(x_1, x_2) * F[\sin \chi(k)]$$

where

$$\sigma = \frac{\pi}{\lambda_0 E}$$

with λ_0 being the electron wavelength, E the acceleration potential, and $\phi(x_1, y_1)$ the projection of the mean inner potential of the object [260]. Therefore quantitative analysis of the image is essential in order to establish its relation to the object.

13.5.1.3 The Damping Function

In view of the 'unusual' transfer functions which are often used in polymer microscopy, the effect of the damping functions $D_1(\boldsymbol{k})$, $D_2(\boldsymbol{k}, \Delta f)$, and $D_3(\boldsymbol{k}, M)$ on the transfer function T must also be considered:

$$T = \exp[i\lambda(\boldsymbol{k}\Delta f)]\, D_1(\boldsymbol{k})\, D_2(\boldsymbol{k}, \Delta f) \times D_3(\boldsymbol{k}, M)\, A_1(\boldsymbol{k})$$

where M is the magnification, D_1 is the damping envelope due to defocus spread, D_2 is the damping envelope due to beam divergence, D_3 is the damping due to the film response, and A_1 is the aperture function [301].

From these equations, the defocus, the maximum defocus spread, and the limiting semi-angle of illumination for any microscope can be determined; these quantities are required to image specific features in the polymer [154, 301, 302]. In practice, the electron dose may be too high at the required magnification.

13.5.1.4 Image Simulation

Polymer microscopists have mainly been using commercially available program packages such as MACTEMPAS (Lawrence Berkeley Laboratory) and CERIUS (Molecular Simulations, Inc.), which all use the multislice approach of Cowley and Moodie [303]. The effect of the contrast transfer function and dynamical scattering on the image are taken into account; therefore, both quantities need to be known quite accurately. In addition, the atomic positions are required. If these are available from an X-ray structure

analysis it is relatively simple to calculate the image for various thicknesses. However, frequently in polymer science, especially in the case of new, innovative materials, these are not available and one of the methods described in Sec. 13.4 must be applied.

A possible way of proceeding is as follows. The microscope transfer function is monitored by a computer with fast Fourier transform capabilities attached to the electron microscope via a suitable CCD camera. On the basis of the electron diffraction pattern from the sample, a decision is made as to which spatial frequencies are to be transferred, and the appropriate phase contrast transfer function chosen [153, 154, 159, 164].

In order to interpret the images, the atomic coordinates are required. If the crystals are too small for X-ray structure analysis and refinement, these parameters must be determined from electron diffraction [159, 163–165, 215–246]. A series of electron diffraction patterns showing different projections is required. Using these, a model structure is proposed, which eventually has to satisfy all the following criteria: (1) all intensities from all projections must give a good agreement with experiment; (2) the calculated density must agree with experiment; (3) the packing energy must be negative; and (4) there must be no nonallowed bond distances or bond angles; (5) possible hydrogen bond patterns must be considered. As indicated in Sec. 4.2 of this Chapter this step is extremely demanding.

The coordinates of all the atoms in the unit cell can be obtained from the model processed in the HRTEM mode of a suitable modelling program to calculate the image convoluted by the correct transfer function and taking account of the sample thickness. High resolution images and their calculated models can thus be obtained and compared with one another. As slow-scan CCD cameras with increased resolution and at lower cost become available, they will undoubtedly play an increasingly important role in high resolution imaging of polymers, because the correct transfer function can be found very quickly from the Fourier Transform of the image.

As an alternative to TEM, STM and AFM will play an increasingly important role in high resolution imaging, as discussed in appropriate preceding sections of this Chapter. However, their limitations should also be kept in mind: (a) it is very difficult to obtain the same high quality images from polymers as those which have been achieved for monomers; and (b) the substrate controls the molecular packing, which is expected to be different from that in crystals, thin films, or bulk material.

13.5.2 Experimental High Resolution Images of Polymers and their Monomeric Analogs

The first high resolution images of poly (p-phenylene terephthalamide) fibers [304] clearly showed the lateral packing and intermolecular registry between PPTA molecules and confirmed their extended chain structure. The size and shape of crystalline domains in PPTA as well as their relative orientations were analyzed and compared with the results of X-ray diffraction experiments [305].

When image restoration methods became generally available, a whole series

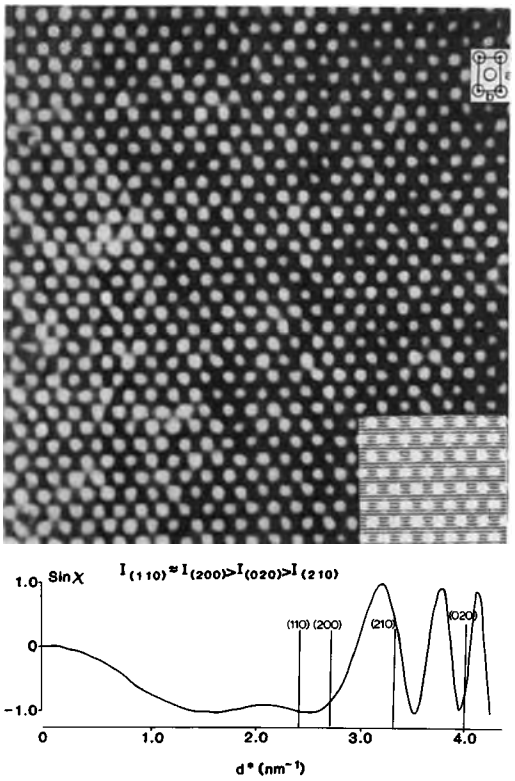


Figure 55. High resolution, low dose image of a monolamellar *n*-paraffin (n-$C_{44}H_{90}$) crystal obtained with a cryoelectron microscope. The image was averaged by correlation techniques. The inset shows a multislice *n*-beam dynamical calculation carried out for the known crystal structure and instrumental conditions used to obtain the micrograph [310].

of polymers were investigated using high resolution TEM methods: isotactic polystyrene [306], polydiacetylene [307], polytetrafluoroethylene [308], poly(phenylene sulfide) [309], and single crystals of alkanes [310] (Fig. 55), polyethylene [311, 312], and polyparaxylylene [313]. Lattice images from poly(paraphenylene benzobisthiazole) and poly(paraphenylene benzobisoxazole) showed specific defects within crystals, and indicated local bending of molecules within compressive failure zones [314].

These observations were used to develop a classification scheme for grain boundary defects in extended-chain polymers [314]. In syndiotactic polystyrene, antiphase boundaries, due to local variations in the crystallographic packing of the helical sPS chains, were revealed [315]. An investigation of polyimides provided evidence for dislocation-mediated splay distortions of lamellae [316]. Lattice images have also been obtained from biopolymers such as cellulose [317, 318] and bacterially synthesized poly(ester)poly(hydroxybutyrate) [319].

From high resolution images of low molar mass organic molecules, such as paraffins [17] and Langmuir–Blodgett films of phthalocyanines [244, 245], it has been possible to characterize the packing of molecular chains and the defects at grain boundaries. In particular, it was possible to characterize the high concentration of defects and misorientation in thin monolayer films. Subsequently, the structure of lipid systems exhibiting lamellar textures were studied and crystal boundaries and local fluctuations of lamellar orientation investigated [320].

The defects in a series of polymeric liquid crystals with highly functionalized mesogenic groups, and their monomeric analogs, were analyzed using cross-correlation methods and their crystal structures analyzed and refined [159, 160]. Recently, a series of such molecules was imaged using high resolution TEM [165, 167, 168, 246], and their atomic positions were determined by simulation of diffraction patterns and maximum entropy crystallographic methods [167, 168, 246]. Using these atomic positions, the images were calculated and compared with the high resolution TEM images [163, 164, 167, 168] (Fig. 56).

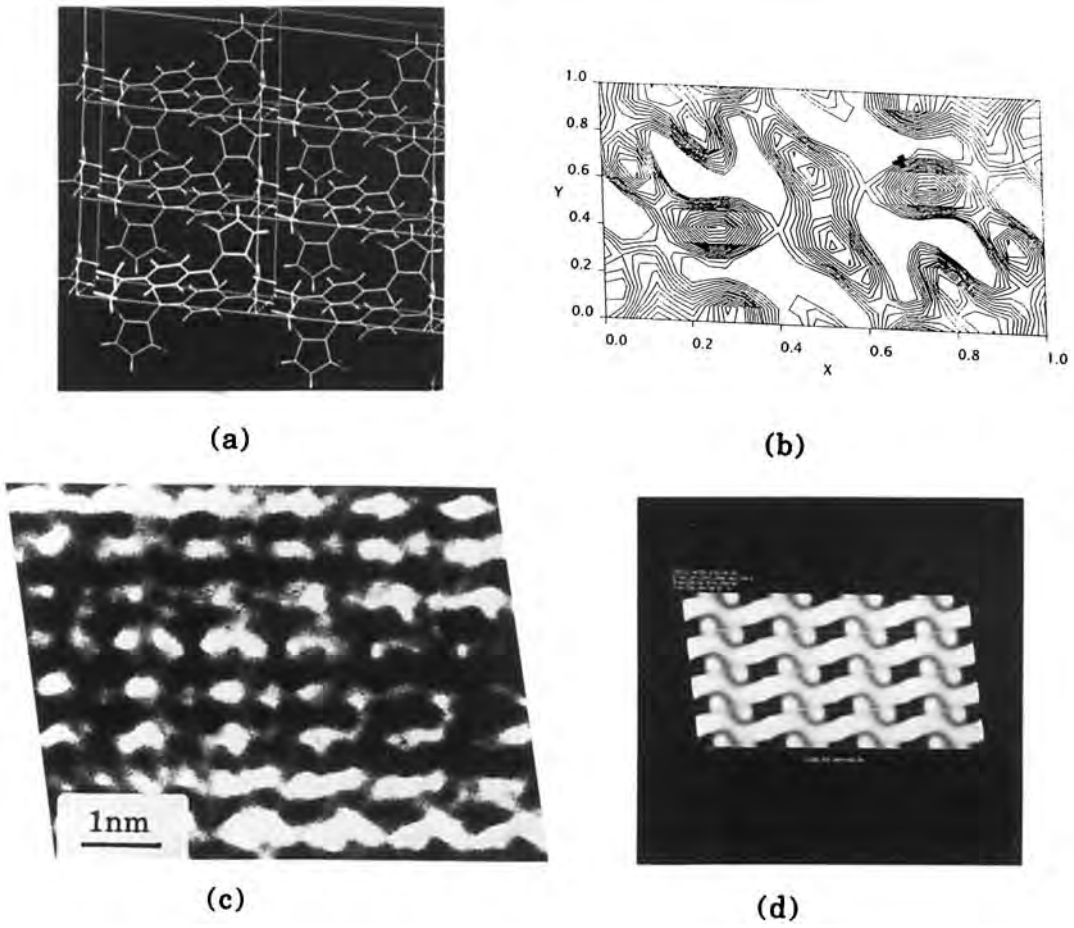


Figure 56. Comparison between the model structure (a), potential map (b), high resolution image (c), and simulated image (d) [247].

Figure 56 shows the dione molecule which was investigated for possible NLO applications. A series of images at different defocus and thickness values is shown in Fig. 57.

From this long list of fairly recent successful high resolution imaging in polymers and their monomeric analogs, it is apparent that the method, as well as the interpretation of images and defects, is becoming increasingly sophisticated.

13.6 Summary

As indicated in the Introduction, electron microscopy and diffraction have played a major role in the investigation of polymers. Some typical applications have been summarized in this Chapter. In writing this article, a possible route could have been to produce a long list of successful investigations on a wide variety of different polymers. However, this would have detracted

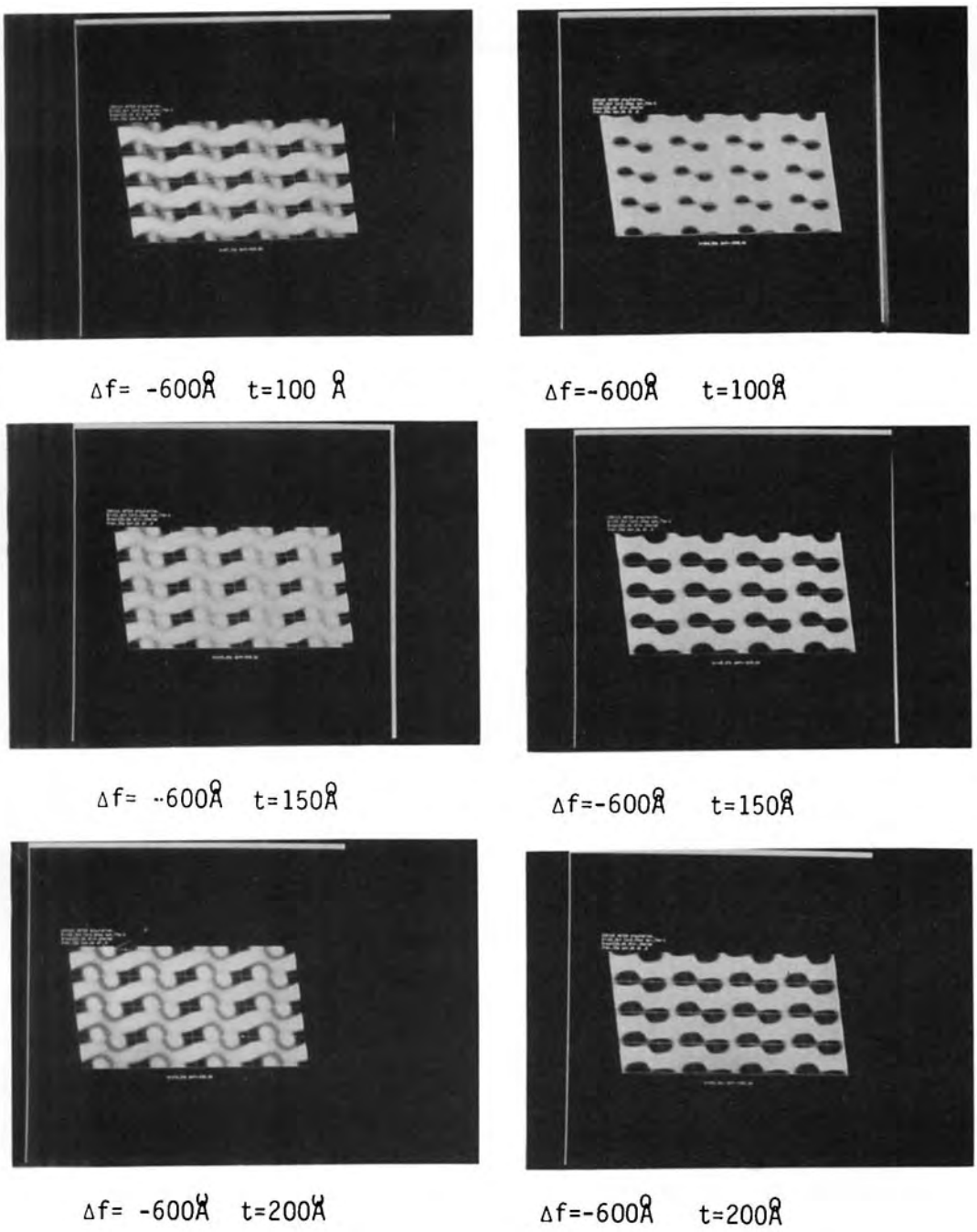


Figure 57. Simulated images obtained from model structures for different defocus (Δf) and thickness (t) values.

from some very exciting new developments currently taking place. Sophisticated methods of analyzing defects have been developed and detailed information can be obtained at atomic level, even from organic samples, by using a combination of electron crystallography and high resolution imaging or by applying one of the new near-field methods. It has been indicated why this information is needed to produce specific physical properties and the kind of molecular architecture which is required. It has also been shown that the difficulties associated with electron crystallography and the interpretation of the high resolution images from organic samples can be overcome. Frequently the situation arises that theoretical papers overemphasize the problems associated with the interpretation of diffraction patterns and images, so that experimentalists are discouraged from even attempting to use the technique. Hopefully the present Chapter will encourage more experimentalists in the polymer field to continue their struggle against the, admittedly awesome, difficulties and, if necessary, to collaborate with a suitable theoretician if the results are difficult to interpret.

Acknowledgements

I am extremely grateful for the immediate and generous response of my colleagues, who supplied figures and reprints of their recent publications.

13.7 References

[1] P. J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY **1967**.
[2] B. Wunderlich, *Macromolecular Physics*, Academic Press, New York **1973**, Vol. 1.
[3] F. Khoury, E. Passaglia in *Treatise on Solid State Chemistry* (Ed.: B. Mannay), Plenum Press, New York **1976**, Vol. 3.
[4] I. G. Voigt-Martin, L. Mandelkern in *Handbook of Polymer Science and Technology* (Ed.: N. Chereremisinoff), M. Dekker, New York **1989**.
[5] W. D. Niegisch, P. R. Swan, *J. Appl. Phys.* **1960**, *31*, 1906.
[6] D. H. Reneker, P. H. Geil, *J. Appl. Phys.* **1960**, *31*, 1916.
[7] D. C. Bassett, F. C. Frank, A. Keller, *Nature (Lond.)* **1959**, *184*, 810.
[8] D. C. Bassett, A. Keller, *Phil. Mag.* **1961**, *6*, 345.
[9] J. B. Jackson, P. J. Flory, R. Chiang, *Trans. Faraday Soc.* **1963**, *59*, 1906.
[10] R. K. Sharma, L. Mandelkern, *Macromolecules* **1969**, *2*, 266.
[11] E. W. Fischer, R. Lorentz, *Kolloid Z.Z. Polym.* **1963**, *189*, 97.
[12] L. Mandelkern, A. J. Peacock, *Polym. Bull.* **1986**, *16*, 529.
[13] S. Krimm, J. M. C. Ching, *Macromolecules* **1972**, *5*, 209.
[14] T. Williams, J. D. Blundell, A. Keller, I. Ward, *J. Polym. Sci. A2* **1968**, *6*, 1613.
[15] D. Y. Yoon, P. J. Flory, *Disc. Faraday Soc.* **1979**, 68.
[16] P. J. Flory, D. Y. Yoon, K. A. Dill, *Macromolecules* **1984**, *17*, 862.
[17] D. L. Dorset, J. Hanlon, C. H. McConnell, J. R. Fryer, B. Lotz, J. C. Wittmann, E. Beckmann, F. Zemlin, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 1696.
[18] D. Dorset, R. Alamo, L. Mandelkern, *Macromolecules* **1993**, *26*, 3143.
[19] A. I. Kitaigorodskii, *Organic Chemical Crystallography*, Consultants Bureau, New York **1961**.
[20] A. Lovinger, D. D. Davis, *Macromolecules* **1986**, *19*, 1861.
[21] J. Boon, E. P. Magré, *Macromol. Chem.* **1969**, *126*, 130.
[22] B. J. Tabor, E. P. Magré, J. Boon, *Eur. Pol. J.* **1971**, *7*, 1127.
[23] A. Lovinger, *Macromolecules* **1986**, *19*, 1861.
[24] A. Lovinger, *J. Polym. Sci. (Phys.)* **1983**, *21*, 97.
[25] B. Lotz, J. C. Wittmann, *J. Polym. Sci. (Phys.)* **1986**, *24*, 1541.
[26] S. Brückner, S. Meille, *Nature* **1989**, *340*, 455.
[27] A. Lovinger, B. Lotz, D. Davis, F. I. Padden Jr, *Macromolecules* **1993**, *26*, 3494.
[28] A. Lovinger, B. Lotz, D. Davis, M. Schumacher, *Macromolecules* **1994**, *27*, 6603.
[29] D. W. Van Krevelen, *Properties of Polymers*, Elsevier, Amsterdam **1976**.

[30] L. Mandelkern, *Crystallisation of Polymers*, McGraw-Hill, New York, **1964**.
[31] G. M. Stack, I. G. Voigt-Martin, L. Mandelkern, *Polym. Bull.* **1982,** *8*, 421.
[32] W. B. Hillig, *Acta Metall.* **1966,** *14*, 1868.
[33] P. D. Calvert, D. R. Uhlman, *J. Appl. Phys.* **1972,** *43*, 944.
[34] I. C. Sanchez, E. A. di Marzio, *Macromolecules* **1971,** *4*, 677.
[35] J. I. Lauritzen, *J. Appl. Phys.* **1973,** *44*, 4353.
[36] J. Maxfield, L. Mandelkern, *Macromolecules* **1977,** *10*, 1141.
[37] D. C. Bassett, A. M. Hodge, R. H. Olley, *Proc. R. Soc. London, Ser. A* **1981,** *377*, 39.
[38] R. H. Olley, D. C. Bassett, D. J. Blundell, *Polymer* **1986,** *27*, 344.
[39] D. C. Bassett, *Phil. Trans. R. Soc. London, Ser. A* **1994,** *348*, 29.
[40] D. R. Norton, A. Keller, *Polymer* **1985,** *26*, 704.
[41] A. Lovinger, D. Davis, *J. Appl. Phys.* **1985,** *58*, 2843.
[42] E. W. Fischer, *Polym. J.* **1984,** *17*, 307.
[43] J. D. Hoffman, *Polymer* **1982,** *23*, 656.
[44] J. D. Hoffman, L. J. Frolen, G. S. Ross, J. I. Lauritzen Jr, *J. Res. Natl. Bur. Stand.* **1975,** *79A*, 671.
[45] L. Mandelkern, N. L. Jain, H. Kim, *J. Polym. Sci.* **1968,** *A-2*, 165.
[46] E. Ergoz, J. G. Fatou, L. Mandelkern, *Macromolecules* **1972,** *5*, 147.
[47] R. Alamo, J. G. Gaton, J. Guzman, *Polymer* **1982,** *23*, 379.
[48] L. Mandelkern, M. Glotin, R. S. Benson, *Macromolecules* **1981,** *14*, 22.
[49] G. Kaning, *Prog. Colloid Polym. Sci.* **1975,** *57*, 176.
[50] I. G. Voigt-Martin, E. W. Fischer, L. Mandelkern, *J. Polym. Sci. (Phys.)* **1980,** *18*, 2347.
[51] I. G. Voigt-Martin, L. Mandelkern, *J. Polym. Sci. (Phys.)* **1984,** *22*, 1901.
[52] I. G. Voigt-Martin in *Advances in Polymer Science* (Eds.: H. H. Kausch, H. G. Zachmann), Springer Verlag, Berlin **1985**, p. 195.
[53] G. M. Stack, I. G. Voigt-Martin, L. Mandelkern, *Macromolecules* **1984,** *17*, 321.
[54] G. Strobl, M. Schneider, I. G. Voigt-Martin, *J. Polym. Sci. (Phys.)* **1980,** *18*, 1361.
[55] I. G. Voigt-Martin, *J. Polym. Sci. (Phys.)* **1980,** *18*, 1513.
[56] I. G. Voigt-Martin, A. J. Peacock, L. Mandelkern, *J. Polym. Sci.* **1989,** *27*, 957.
[57] I. G. Voigt-Martin, L. Mandelkern, *J. Polym. Sci.* **1989,** *27*, 967.
[58] D. C. Martin in *Trends in Polymer Science*, Elsevier, Cambridge **1992**.
[59] D. H. Reneker, J. Mazur, *Polymer* **1983,** *24*, 1387.
[60] D. Dorset, *Macromolecules* **1993,** *26*, 3146.
[61] P. M. Wilson, D. C. Martini, *J. Mat. Res.* **1992,** *7*, 3150.
[62] K. J. Ihn, J. A. N. Zasadzinski, R. Pindak, A. J. Slaney, J. Goodby, *Science* **1992,** *258*, 275.
[63] S. Hudson, E. L. Thomas, *Phys. Rev. Lett.* **1989,** *62*, 1993.
[64] I. G. Voigt-Martin, R. W. Garbella, U. Valleriien, *J. Phys. (France)* **1992,** *2*, 345.
[65] D. C. Martin, E. L. Thomas, *Phil. Mag.* **1991,** *A64*, 903.
[66] A. Noshay, J. E. McGrath, *Block Co-polymers: Overview and Comprehensive Treatise*, Academic Press, New York **1977**.
[67] R. I. Ceresa (Ed.), *Block and Craft Copolymerisation*, Wiley, New York **1973**.
[68] E. Helfand, *Macromolecules* **1975,** *8*, 552.
[69] L. Leibler, *Macromolecules* **1980,** *13*, 1602.
[70] H. Fried, K. Binder, *J. Chem. Phys.* **1991,** *94*, 8349.
[71] F. S. Bates, G. H. Frederickson, *Ann. Rev. Phys. Chem.* **1990,** *41*, 525.
[72] E. L. Thomas, D. Alward, D. J. Kinning, D. C. Martini, D. Handlin, L. J. Fetters, *Macromolecules* **1986,** *19*, 2197.
[73] P. D. Olmsted, S. T. Milner, *Phys. Rev. Lett.* **1994,** *72*, 936.
[74] E. L. Thomas, D. M. Anderson, C. S. Henkee, D. Hoffman, *Nature* **1988,** *334*, 598.
[75] M. M. Disko, K. S. Liang, S. K. Behal, R. J. Roe, K. J. Jein, *Macromolecules* **1993,** *26*, 2983.
[76] I. W. Hamley, K. A. Koppi, I. H. Rosedale, F. S. Bates, K. Almdal, K. Mortensen, *Macromolecules* **1993,** *26*, 5959.
[77] D. A. Hajduk, P. E. Harper, S. M. Gruner, C. Honeker, G. Kim, E. L. Thomas, L. J. Fetters, *Macromolecules* **1994,** *27*, 4063.
[78] L. E. Scriven, *Nature* **1976,** *263*, 123.
[79] D. M. Anderson, H. T. Davis, L. E. Scriven, J. C. C. Nitsche, *Adv. Chem. Phys.* **1990,** *37*, 337.
[80] D. M. Anderson, E. L. Thomas, *Macromolecules* **1988,** *21*, 3221.
[81] R. Stadler, C. Auschra, J. Beckmann, U. Krappe, I. G. Voigt-Martin, L. Leibler, *Macromolecules* **1995,** *28*, 3080.
[82] U. Krappe, R. Stadler, I. G. Voigt-Martin, *Macromolecules* **1995,** *28*, 4558.
[83] C. Auschra, R. Stadler, I. G. Voigt-Martin, *Polymers* **1992,** *34*, 2082.
[84] C. Auschra, R. Stadler, *Macromolecules* **1993,** *26*, 6364.
[85] S. P. Gido, J. Gunther, E. L. Thomas, D. Hoffman, *Macromolecules* **1993,** *26*, 2636.
[86] H. Hasegawa, H. Tanaka, K. Yamasaki, T. Hashimoto, *Macromolecules* **1987,** *20*, 1651.

[87] E. Helfand, Z. R. Wasserman, *Macromolecules* **1978,** *11*, 960.
[88] E. Helfand, Z. R. Wasserman, *Macromolecules* **1978,** *11*, 960.
[89] K. Kawasaki, T. Ohta, *Makromolecules* **1986,** *19*, 2621.
[90] L. Leibler, *Makromolecules* **1980,** *13*, 1602.
[91] K. Binder, G. H. Frederickson, *J. Chem. Phys.* **1989,** *91*, 7265.
[92] I. D. Verhoeven, *Fundamentals of Physical Metallurgy*, Wiley, New York **1975**.
[93] J. P. Hirth, J. Lothe, *Theory of Dislocations*, 2nd edn, Wiley, New York **1982**.
[94] H. F. Scherk, *J. Reine Angew. Math.* **1835**, *13*, 185.
[95] E. E. Underwood, *Quantitative Stereology*, Addison-Wesley, Reading, MA **1970**.
[96] O. Olabisi, L. M. Robeson, M. T. Shaw, *Polymer—Polymer Miscibility*, Academic Press, London **1979**.
[97] H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena*, Oxford Science Publishers, Oxford **1971**.
[98] A. Sariban, K. Binder *Macromolecules* **1988,** *21*, 711.
[99] K. Binder, *Colloid Polym. Sci.* **1988,** *266*, 871.
[100] K. W. Kehr, K. Binder, in *Applications of the Monte Carlo Method in Statistical Physics* (Ed.: K. Binder), Springer Verlag, New York **1987,** p. 181.
[101] V. L. Ginsburg, *Sov. Phys. Solid State* **1960,** *1*, 1824.
[102] P. G. de Gennes, *J. Phys. Lett. (Paris)* **1977,** *38*, L441.
[103] K. Binder, *Phys. Rev., Ser. A* **1984,** *29*, 341.
[104] I. E. G. Lipson, *Macromolecules* **1991,** *24*, 1334.
[105] F. J. Guevara-Rodriguez, F. Aquilera-Granja, R. Kikuchi, *J. Phys. (Paris)* **1994,** *4*, 589.
[106] M. Lifschitz, I. Dudowicz, K. F. Freed, *J. Chem. Phys.* **1994,** *100*, 3957.
[107] K. S. Schweizer, *Macromolecules* **1993,** *26*, 6033.
[108] K. S. Schweizer, *Macromolecules* **1993,** *26*, 6050.
[109] M. D. Gehlen, J. H. Rosedale, F. S. Bates, G. D. Wignall, K. Almdal, *Phys. Rev. Lett.* **1992,** *68*, 2452.
[110] H. Ito, T. P. Russel, G. Wignall, *Macromolecules* **1987,** *20*, 2213.
[111] H. M. Shabana, W. Guo, R. H. Olley, D. C. Bassett, *Polymer* **1993,** *34*, 1313.
[112] L. P. McMaster, *Macromolecules* **1973,** *6*, 760.
[113] I. G. Voigt-Martin, K. H. Leister, R. Rosenau, R. Koningsvelo, *J. Polym. Sci. (Phys.)* **1986,** *24*, 723.
[114] R. Koningsveld, H. A. G. Chermin, M. Gordon, *Proc. R. Soc., London, Ser. A* **1970,** *319*, 331.
[115] J. W. Cahn, *Trans. Met. Soc. AIME* **1968,** *242*, 166.
[116] T. Nishi, T. Wang, T. Kwei, *Macromolecules* **1975,** *8*, 227.
[117] Y. Chou, W. Goldburg, *Phys. Rev.* **1978,** *69*, 725.
[118] T. Ohta, K. Kawasaki, *Phys. Lett.* **1978,** *64A*, 404.
[119] E. D. Siggia, *Phys. Rev., Ser. A* **1979,** *20*, 595.
[120] B. Kalb, A. J. Pennings, *J. Mater. Sci.* **1980,** *15*, 2584.
[121] G. Capacio, A. G. Gibson, I. M. Ward in *Ultra-High Modulus Polymers* (Eds.: A. Ciferri, I. M. Ward), Applied Science, London **1979**.
[122] A. E. Zachariades, W. T. Mead, R. S. Porter in *Ultra-High Modulus Polymers* (Eds.: A. Ciferri, I. M. Ward), Applied Science, London **1979**.
[123] G. Capaccio, A. Crompton, I. M. Ward, *J. Polym. Sci. (Phys.)* **1976,** *14*, 164.
[124] E. S. Clark, L. S. Scott, *Polym. Eng. Sci.* **1974,** *14*, 682.
[125] M. Miles, J. Petermann, H. Gleiter, *Prog. Colloid Polym. Sci.* **1977,** *62*, 6.
[126] R. H. Olley, D. C. Bassett, P. J. Hine, I. M. Ward, *J. Mater. Sci.* **1993,** *28*, 1107.
[127] W. Wade Adams, D. Yang, E. L. Thomas, *J. Mater. Sci.* **1986,** *21*, 2239.
[128] M. G. Dobb, I. E. McIntyre, *Adv. Polym. Sci.* **1984,** *60/61*.
[129] E. J. Roche, T. Takahashi, E. L. Thomas in *Fibre Diffraction Methods* (Eds.: A. D. French, K. H. Gardner), A.C.S. Symp. Ser. No. 141, American Chemical Society, Washington, DC **1980**.
[130] S. R. Allen, A. G. Fillippov, R. J. Farris, E. L. Thomas in *The Strength and Stiffness of Polymers* (Eds.: A. Zachariades, R. S. Porter), Marcel Dekker, New York **1984**.
[131] L. Sawyer, M. Jaffe, *J. Mater. Sci.* **1986,** *21*, 360.
[132] J. B. Stamatoff, *Mol. Cryst. Liq. Cryst.* **1984,** *110*, 75.
[133] P. G. de Gennes, *The Physics of Liquid Crystals*, Clarendon Press, Oxford **1971**.
[134] S. Chandresekhar, *Liquid Crystals*, Cambridge University Press, Cambridge **1977**.
[135] H. Ringsdorf, I. G. Voigt-Martin, J. Wendorff, R. Wüstefeld, R. Zentel in *Chemistry and Physics of Macromolecules* (Eds.: E. W. Fischer, H. Sillescu), Deutsche Forschungsgemeinschaft VCH Verlagsgesellschaft **1991** Weinheim.
[136] P. G. Flory, G. Ronca, *Mol. Cryst. Liq. Cryst.* **1979,** *28*, 289.
[137] J. Nehring, A. Saupe, *J. Chem. Phys.* **1972,** *56*, 5527.

[138] W. L. McMillan, *Phys. Rev., Ser. A* **1987,** *4*, 236.
[139] A. M. Donald, A. H. Windle, *J. Mater. Sci.* **1983,** *18*, 1143.
[140] A. M. Donald, A. H. Windle, *J. Mater. Sci.* **1984,** *19*, 2095.
[141] A. M. Donald, A. H. Windle, *Polymer* **1983,** *25*, 1235.
[142] E. L. Thomas, B. Wood, *Faraday Disc. Chem. Soc.* **1985,** *79*, 15.
[143] S. D. Hudson, A. J. Lovinger, R. G. Larson, D. D. Davis, R. O. Garay, K. Fujishoro, *Macromolecules* **1993,** *26*, 5643.
[144] D. Demus, L. Richter, *Textures in Liquid Crystals*, Verlag Chemie, Weinheim **1978.**
[145] F. C. Frank, *Phil. Mag.* **1958,** *42*, 1014.
[146] S. Chandresekhar, G. S. Ranganath, *Adv. Phys.* **1986,** *35*, 507.
[147] H. Finkelmann, *Angew. Chem., Int. Ed. Engl.* **1987,** *26*, 816.
[148] D. S. Chemla, J. Zyss (Eds.), *Non-linear Optical Properties of Organic Molecules and Crystals*, Academic Press, London **1987,** Vol. 1.
[149] I. G. Voigt-Martin, E. Zhou, P. Simon, R. W. Garbella, D. Yan, W. Paulus, H. Ringsdorf, *Adv. Mater. Opt. Electron.* **1993,** *2*, 245.
[150] G. Taylor (Ed.), *Ferroelectric Liquid Crystals*, Gordon & Breach, London **1991.**
[151] I. G. Voigt-Martin, H. Durst, *Rapid Commun.* **1986.**
[152] I. G. Voigt-Martin, H. Durst, *Liquid Cryst.* **1987,** *2*, 601.
[153] I. G. Voigt-Martin, H. Durst, B. Reck, H. Ringsdorf, *Macromolecules* **1988,** *21*, 1620.
[154] I. G. Voigt-Martin, H. Krug, D. Van Dyck, *J. Phys. (France)* **1990,** *51*, 2347.
[155] Landau and Lifschitz, *Statistical Physics*, Pergamon, London **1959.**
[156] J. Als Nielsen, J. D. Litster, R. Birgenau, M. Kaplan, C. R. Safinya, A. Lindegaard-Andersen, S. Mathiesen, *Phys. Rev., Ser. B* **1980,** *22*, 312.
[157] I. G. Voigt-Martin in *Crystallisation of Polymers* (Ed.: M. Dosière), NATO ASI Series, Kluwer Academic, New York **1994.**
[158] J. Selinger, R. Bruinsma, *Phys. Rev., Ser. A* **1991,** *43*, 2910.
[159] I. G. Voigt-Martin, R. W. Garbella, M. Schumacher, *Macromolecules* **1992,** *25*, 961.
[160] I. G. Voigt-Martin, R. W. Garbella, M. Schumacher, *Liquid Crystals* **1994,** *17*, 775.
[161] M. Ballauf, G. F. Schmidt, *Mol. Cryst. Liq. Cryst.* **1987,** *147*, 163.
[162] H. Ringsdorf, P. Tschirner, O. Herrmann-Schönherr, J.-H. Wendorff, *Makromol. Chem.* **1987,** *188*, 1431.
[163] I. G. Voigt-Martin, P. Simon, S. Bauer, H. Ringsdorf, *Macromolecules* **1995,** *28*, 236.
[164] I. G. Voigt-Martin, P. Simon, D. H. Yan, A. Yakimanski, S. Bauer, H. Ringsdorf, *Macromolecules* **1995,** *28*, 243.
[165] I. G. Voigt-Martin, *Microsc. Soc. Am. Bull.* **1995,** *24*, 527.
[166] D. Adams, F. Closs, D. Funhoff, D. Haarer, H. Ringsdorf, P. Schumacher, K. Siemensmeyer, *Phys. Rev. Lett.* **1993,** *70*, 457.
[167] I. G. Voigt-Martin, D. H. Yan, C. J. Gilmore, K. Shankland, G. Bricogne, *Ultramicroscopy* **1994,** *56*, 271.
[168] I. G. Voigt-Martin, P. Simon, A. Yakimanski, *Ultramicroscopy* (in press).
[169] W. Stocker, S. N. Magonov, H. J. Cantow, J. C. Wittman, B. Lotz, *Macromolecules* **1993,** *26*, 5915.
[170] B. Lotz, A. Lovinger, R. W. Cais, *Macromolecules* **1988,** *21*, 2375.
[171] W. Stocker, G. Bar, M. Kunz, M. Möller, S. N. Magonov, H. J. Cantow, *Polym. Bull.* **1991,** *26*, 215.
[172] H. Hausma, F. Motademi, P. Smith, P. Hansana, J. C. Wittman, *Polymer* **1992,** *33*, 647.
[173] S. N. Magonov, H. J. Cantow, *J. Appl. Polym. Sci. Appl. Polym. Symp.* **1992,** *51*, 3.
[174] D. Snetivy, G. J. Vansco, *Macromolecules* **1992,** *25*, 3320.
[175] W. Stocker, M. Schumacher, S. Graff, J. Lang, J. C. Wittman, A. J. Lovinger, B. Lotz, *Macromolecules* **1994,** *27*, 6948.
[176] S. Chiang in *Scanning Tunnel Microscopy I* (Eds.: H. J. Güntherodt, R. Wiesendanger), Springer, Berlin **1992.**
[177] J. P. Rabe, *Ultramicroscopy* **1992,** *41*, 42.
[178] H. J. Butt, R. Guckenberger, J. P. Rabe, *Ultramicroscopy* **1992,** *46*, 375.
[179] E. Stoll, *IBM J. Res.* **1991,** *35*, 67.
[180] D. M. Eigler, P. S. Weiss, E. K. Schweitzer, N. D. Lang, *Phys. Rev. Lett.* **1991,** *66*, 1189.
[181] E. Stoll, *IBM J. Res.* **1991,** *35*, 67.
[182] G. Binning, H. Rohrer, *Rev. Mod. Phys.* **1987,** *59*, 615.
[183] H. Ohtani, R. J. Wilson, S. Chiang, C. M. Mate, *Phys. Rev. Lett.* **1988,** *60*, 2398.
[184] D. Tomanek, S. G. Lonie, H. J. Mamin, D. W. Abraham, R. E. Thomson, E. Ganz, J. Clarke, *Phys. Rev., Ser. B* **1987,** *37*, 7790.
[185] M. Hara, Y. Iwakabe, K. Tochigi, H. Sasabe, A. F. Garito, A. Yamada, *Nature* **1990,** *344*, 228.
[186] J. S. Foster, J. E. Frommer, *Nature* **1988,** *333*, 542.
[187] J. K. Spong, H. A. Mizes, L. J. Lacomb, M. M. Dovek, J. E. Frommer, J. S. Foster, *Nature* **1989,** *338*, 137.
[188] D. P. E. Smith, J. K. H. Hörber, G. Binnig, H. Nejoh, *Nature* **1990,** *344*, 641.

[189] G. C. McGonigal, R. H. Bernhardt, D. J. Thomson, *Appl. Phys. Lett.* **1990,** *57*, 28.
[190] J. P. Rabe, S. Buchholz, *Science* **1991,** *253*, 424.
[191] J. P. Rabe, S. Buchholz, *Macromol. Chem. Macromol. Symp.* **1991,** *50*, 261.
[192] J. P. Rabe in *Nanostructures Based on Molecular Materials* (Eds.: W. Göpel, C. Ziegler), VCH, Weinheim **1992.**
[193] T. Sleator, R. Tycko, *Phys. Rev. Lett.* **1988,** *60*, 1418.
[194] J. P. Rabe, S. Buchholz in *Conjugated Polymeric Materials: Opportunities in Electronics, Optoelectronics and Molecular Electronics* (Eds.: J. L. Bredas, R. R. Chance), NATO-ARW Series E, Kluwer, Dordrecht **1990.**
[195] S. N. Magonov, J. Schuchhardt, S. Kempf, E. Keller, H. J. Cantow, *Synth. Methods* **1991,** *40*, 59.
[196] J. P. Rabe, M. Sano, D. Batchelder, A. A. Kolatchev, *J. Microsc.* **1988,** *152*, 573.
[197] T. J. McMaster, H. J. Carr, M. J. Miles, P. Cairns, V. J. Morris, *Macromolecules* **1991,** *24*, 1428.
[198] M. Amrein, R. Dürr, A. Stasiak, H. Gross, G. Travaglini, *Science* **1989,** *243*, 1708.
[199] R. Guckenberger, W. Wiegräbe, A. Hillebrand, T. Hartmann, Z. Wang, W. Baumeister, *Ultramicroscopy* **1989,** *31*, 327.
[200] D. M. Huong, M. Drechsler, M. Möller, H. J. Cantow, *J. Microsc.* **1992,** *166*, 317.
[201] H. J. Cantow, M. Kuntz, S. Klotz, M. Möller, *Macromol. Chem. Makromol. Symp.* **1989,** *23*, 57.
[202] A. Du Chesne, G. Lieser, G. Wegner, *Colloid Polym. Sci.* **1994,** *272*, 1329.
[203] M. Kunz, M. Möller, H. J. Cantow, *Macromol. Chem.* **1987,** *8*, 401.
[204] R. Castening, C. R. Henry, *Seances Acad. Sci., Ser. B* **1992,** *255*, 76.
[205] E. L. Thomas, D. B. Alward, D. J. Kinning, D. C. Martin, D. L. Handlin, L. J. Fetters, *Macromolecules* **1986,** *19*, 2197.
[206] D. L. Vezie, W. Wade Adams, E. L. Thomas, *Polymer* **1995,** *36*, 1761.
[207] E. A. Burke, *IEEE Trans. Nucl. Sci.* **1980,** *NS-27*(6), 1760.
[208] J. H. Butler, D. C. Toy, G. F. Bradley, S. J. Krause, G. M. Brown, *Microscopy: The Key Research Tool* **1992,** pp. 103–106.
[209] L. Sawyer, D. T. Grubb, *Polymer Microscopy*, Chapman & Hall, New York **1987.**
[210] G. H. Stout, L. H. Jensen, *X-ray Structure Determination*, MacMillan New York **1968.**
[211] D. L. Dorset, *Chem. Phys. Lipids* **1977,** *13*, 20.
[212] E. Day, J. B. Lando, *Macromolecules* **1980,** *13*, 1483.
[213] K. Shimamura, F. E. Karasz, J. A. Hirsch, J. C. W. Chien, *Macromol. Sci.* **1981,** *2*, 473.
[214] H. Hasegawa, W. Claffey, P. H. Geil, *J. Macromol. Sci. Phys.* **1977,** *B13*, 89.
[215] B. Moss, D. L. Dorset, *J. Macromol. Sci. Phys.* **1983,** *B22*, 69.
[216] B. Moss, D. L. Dorset, *Acta Crystallogr., Sect. A* **1983,** *39*, 609.
[217] W. Claffey, K. Gardner, J. Blackwell, J. Lando, P. H. Geil, *Phil. Mag.* **1974,** *30*, 1223.
[218] E. Roche, H. Chanzy, M. Boudeulle, R. H. Marchessault, P. Sundarajan, *Macromolecules* **1978,** *11*, 86.
[219] S. Poulin-Dandurand, S. Perez, J. F. Revol, F. Brisse, *Polymer* **1974,** *20*, 419.
[220] F. Brisse, B. Rémillard, H. Chanzy, *Macromolecules* **1984,** *17*, 1980.
[221] M. Thakur, J. B. Lando, *Macromolecules* **1983,** *16*, 143.
[222] L. I. Tatarinova, B. K. Vainshtain, *Visokomolek. Soed.* **1962,** *4*, 261.
[223] B. K. Vainshtain, L. I. Tatarinova in *Confirmation of Biopolymers* (Ed.: G. Ramachandran), Academic Press, New York **1967,** Vol. 2.
[224] C. Guizard, H. Chanzy, A. Sarko, *J. Mol. Biol.* **1985,** *183*, 397.
[225] C. Guizard, H. Chanzy, A. Sarko, *Macromolecules* **1984,** *17*, 100.
[226] S. Perez, M. Roux, R. H. Marchessault, *J. Mol. Biol.* **1979,** *129*, 113.
[227] K. Mazeau, H. Chanzy, W. T. Winter, *ACS Polym. Prep.* **1992,** *33*(1), 244.
[228] S. V. Meille, S. Brückner, A. J. Lovinger, F. J. Padden, *Macromolecules* **1994,** *27*, 2615.
[229] H. Hu, D. L. Dorset, *Acta Crystallogr., Part B* **1989,** *45*, 283.
[230] D. L. Dorset, *Macromolecules* **1991,** *24*, 1175.
[231] D. L. Dorset, *Proc. Natl. Acad. Sci. USA* **1991,** *88*, 5599.
[232] H. Hu, D. L. Dorset, *Macromolecules* **1990,** *23*, 4604.
[233] M. Boudeulle, *Cryst. Struct. Commun.* **1975,** *4*, 9.
[234] J. C. Chien, F. E. Karasz, K. Shimamura, *Macromolecules* **1982,** *15*, 1012.
[235] W. Claffey, J. Blackwell, *Biopolymers* **1976,** *15*, 1903.
[236] H. Chanzy, S. Perez, D. P. Miller, G. Paradosi, W. T. Winter, *Macromolecules* **1987,** *20*, 2407.
[237] D. L. Dorset, M. P. McCourt, *Trans. Am. Cryst. Assoc.* **1992,** *28*, 105.
[238] J. Brisson, H. Chanzy, W. T. Miller, *Macromolecules* **1991,** *13*, 31.
[239] F. Brisse, A. Palmer, B. Moss, D. Dorset, W. A. Roughead, D. P. Miller, *Eur. Polym.* **1984,** *20*, 791.

[240] S. V. Meille, S. Brückner, A. J. Lovinger, F. J. Padden, *Polymer* **1989,** *30*, 786.
[241] S. V. Meille, S. Brückner, J. B. Lando, *Polymer* **1989,** *30*, 786.
[242] D. L. Dorset, *Adv. Electron. Electron Phys.* **1994,** *88*, 111.
[243] J. Wittman, B. Lotz, *Progr. Polym. Sci.* **1990,** *15*, 909.
[244] J. R. Fryer, C. M. Conell, R. A. Hann, B. L. Eyres, S. K. Gupty, *Phil. Mag.* **1990,** *B61*, 643.
[245] J. R. Fryer, C. M. Conell, G. H. Grant, R. A. Hann, S. K. Gupty, B. L. Eyres, *Phil. Mag.* **1991,** *B63*, 1193.
[246] I. G. Voigt-Martin, D. H. Yan, C. Gilmore, G. Bricogne, *Acta Crystallogr., Part A* **1995,** *A51*, 849.
[247] I. G. Voigt-Martin, D. H. Yan, K. Shankland, C. Gilmore, C. Bricogne, *Ultramicroscopy* **1994,** *56*, 271.
[248] X. Zou, Y. Sukhared, S. Hovmöller, *Ultramicroscopy* **1993,** *49*, 147.
[249] S. Hovmöller, *Ultramicroscopy* **1992,** *41*, 121.
[250] L. Amos, R. Henderson, P. N. T. Unwin, *Prog. Biophys. Mol. Biol.* **1982,** *39*, 183.
[251] I. G. Voigt-Martin, D. M. Yan, R. Wortman, K. Elich, *Ultramicroscopy* **1995,** *57*, 29.
[252] I. A. Clark, *Handbook of Computational Chemistry*, Wiley, New York **1985**.
[253] H. A. Kurz, J. J. Stewart, K. M. Dieter, *J. Comput. Chem.* **1992,** *1*, 82.
[254] J. S. Dewar, W. Thiel, *J. Am. Chem. Soc.* **1977,** *99*, 4907.
[255] N. Karasawa, W. A. Goddard, *J. Phys. Chem.* **1989,** *93*, 7320.
[256] S. L. Mayo, B. D. Olafson, W. A. Goddard, *J. Phys. Chem.* **1990,** *94*, 8897.
[257] S. Kopp, J. C. Wittman, B. Lotz, *Microsc. Soc. Am. Bull.* **1994,** *24*, 557.
[258] S. Perez, J. Revol, *Microsc. Soc. Am. Bull.* **1993,** *23*, 29.
[259] K. Mazean, W. Winter, M. Chanzy, *Macromolecules* **1994,** *27*, 7606.
[260] J. Cowley, *Diffraction Physics*, North Holland, Amsterdam **1986**.
[261] D. Dorset, G. Hauptman, *Ultramicroscopy* **1976,** *1*, 195.
[262] D. Dorset, *J. Electron Microsc. Technol.* **1985,** *2*, 89.
[263] D. Dorset, *Proc. Natl. Acad. Sci. USA* **1990,** *87*, 8541.
[264] D. Dorset, *Ultramicroscopy* **1991,** *38*, 23.
[265] D. Dorset, *Macromolecules* **1991,** *24*, 1175.
[266] D. Dorset, F. Zemlin, *Ultramicroscopy* **1990,** *33*, 227.
[267] D. Dorset, W. P. Zhang, *J. Electron Microsc. Technol.* **1991,** *18*, 142.
[268] P. N. Unwin, R. Henderson, *J. Mol. Biol.* **1975,** *94*, 425.
[269] D. L. Misell, *Image Analysis, Enhancement and Interpretation*, North Holland, Amsterdam **1978**.
[270] H. A. Hauptman, *Crystal Structure Determination. The Role of Cosine Seminvariants*, Plenum Press, New York **1972**.
[271] G. Giacovazzo, H. C. Monaco, D. Viterbo, F. Scordari, E. Gilli, G. Zanotti, M. Catti, *Fundamentals of Crystallography*, Oxford University Press, Oxford **1992**.
[272] D. Dorset, H. Hauptman, *Ultramicroscopy* **1976,** *1*, 195.
[273] D. Dorset, B. J. Jap, M. H. Ho, R. M. Glaeser, *Acta Crystallogr., Part A* **1979,** *35*, 1001.
[274] B. Moss, D. Dorset, *Acta Crystallogr., Part A* **1982,** *38*, 207.
[275] K. Ishizuka, M. Miyazaki, N. Uyeda, *Acta Crystallogr., Part A* **1982,** *38*, 408.
[276] B. K. Vainshtain *Advances in Structure Research by Diffraction Methods* (Ed.: R. Brill), Interscience, New York, Vol. 1 **1976**.
[277] D. Dorset, M. McCourt, J. R. Fryer, W. Tivol, J. Turner, *Microsc. Soc. Am. Bull.* **1994,** *24*, 398.
[278] D. Dorset, *Microsc. Soc. Am. Bull.* **1994,** *24*, 547.
[279] G. de Titta, J. Edmonds, D. A. Langs, H. Hauptman, *Acta Crystallogr., Part A* **1975,** *31*, 472.
[280] D. Dorset, B. K. Jap, M. H. Ho, R. Glaeser, *Acta Crystallogr., Part A* **1979,** *35*, 1001.
[281] D. Dorset, *Acta Crystallogr., Part B* **1994,** *50*, 201.
[281] C. J. Gilmore, K. Shankland, J. Fryer, *Ultramicroscopy* **1993,** *49*, 132.
[282] C. J. Gilmore, K. Shankland, J. Fryer, *Trans. Am. Cryst. Assoc.* **1992,** *28*, 129.
[283] H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, NJ **1946**.
[284] A. Klug, *Acta Crystallogr., Part A* **1958,** *46*, 515.
[285] H. A. Hauptmann, J. Karle, *Solution of the Phase Problem. 1: The Centrosymmetric Crystal*, Polycrystal Book Service **1953**.
[286] M. E. Daniels, *Ann. Math. Stat.* **1954,** *25*, 631.
[287] G. Bricogne, *Acta Crystallogr., Part A* **1984,** *44*, 410.
[288] C. Bricogne in *Crystallographic Computing. 4: Techniques and New Technologies* (Eds.: N. W. Isaacs, M. R. Taylor), Clarendon Press, Oxford Press, **1988,** pp. 60–79.
[289] G. Bricogne, C. J. Gilmore, *Acta Crystallogr., Part A* **1990,** *46*, 284.

[290] C. J. Gilmore, G. Bricogne in *Crystallographic Computing. 5: From Chemistry to Biology* (Eds.: D. Moras, A. D. Podjarny, J. C. Thierry), Oxford University Press, Oxford **1992,** pp. 298–397.
[291] C. Gilmore, *J. Appl. Crystallogr.* **1984,** *17*, 42.
[292] C. Gilmore, *J. Appl. Crystallogr.* **1988,** *21*, 571.
[293] G. Bricogne, *Acta Crystallogr.* **1993,** *D49*, 37.
[294] G. A. Sim, *Acta Crystallogr.* **1959,** *12*, 813.
[295] W. Dong, T. Baird, J. R. Fryer, C. J. Gilmore, D. D. Micnicol, G. Bricogne, D. J. Smith, M. A. O'Keefe, S. Hovmöller, *Nature* **1992,** *355*, 605.
[296] C. J. Gilmore, K. Shankland, G. Bricogne, *Proc. R. Soc., London* **1993,** *442*, 97.
[297] J. C. M. Spence, *Experimental High Resolution Electron Microscopy*, Oxford University Press, Oxford **1988,** Vol. 2.
[298] S. Amelinckx in *Examining the Sub-Micron World* (Eds.: R. Feder, J. McGowan, D. M. Shinozahi), NATO ASI Series B: Physics, Plenum Press, New York **1986,** Vol. 137.
[299] E. W. Fischer, H. Orth, *Macromol. Chem.* **1965,** *88*, 1881.
[300] I. G. Voigt-Martin, E. H. Andrews, *Proc. R. Soc., London, Ser. A.* **1972,** *327*, 251.
[301] D. Martin, E. L. Thomas, *Polymer* **1995,** *36*, 1743.
[302] I. G. Voigt-Martin, H. Durst, *Macromolecules* **1989,** *22*, 168.
[303] J. M. Cowley, A. F. Moodie, *Acta Crystallogr.* **1957,** *10*, 609.
[304] M. G. Dobb, A. M. Hindeleh, D. J. Johnson, B. P. Saville, *Nature* **1975,** *253*, 189.
[305] S. C. Bennett, M. G. Dobb, D. J. Johnson, R. Murray, B. P. Saville, *High Resolution Studies of Electron Beam Sensitive Polymers, EMAG Proc.* **1975,** p. 329.
[306] M. Tsuji, S. K. Roy, R. St. John Manley, *Polymer* **1984,** *25*, 1573.
[307] P. T. Read, R. J. Young, *J. Mater. Sci. Lett.* **1981,** *16*, 2922.
[308] H. Chanzy, T. Folda, P. Smith, K. Gardner, J. F. Revol, *J. Mater. Sci. Lett.* **1986,** *5*, 1045.
[309] A. Uemura, M. Tsuji, A. Kawaguchi, K. Katayama, *J. Mater. Sci.* **1988,** *23*, 1506.
[310] F. Zemlin, E. Reuber, B. Beckman, E. Zeitler, D. Dorset, *Science* 229.
[311] S. Giorgio, R. Kern, *J. Polym. Sci. (Phys.)* **1984,** *22*, 1931.
[312] J. F. Revol, R. St. John Manley, *J. Mater. Sci. Lett.* **1986,** *5*, 249.
[313] M. Tsuji, S. Isoda, M. Ohara, A. Kawaguchi, K. Katayama, *Polymer* **1982,** *23*, 1568.
[314] D. C. Martin, E. L. Thomas, *Phil. Mag.* **1991,** *A64*, 903.
[315] P. Pradere, E. L. Thomas, *Macromolecules* **1990,** *23*, 4954.
[316] J. R. Ojeda, D. Martin, *Macromolecules* **1993,** *26*, 6557.
[317] J. F. Revol, *J. Mater. Sci. Lett.* **1985,** *4*, 1347.
[318] S. Kuga, R. M. Brown, *Polym. Commun.* **1987,** *28*, 311.
[319] J. Revol, H. D. Chanzy, Y. Deslandes, R. H. Marchessault, *Polymer* **1990,** *30*, 1973.
[320] J. R. Fryer, D. L. Dorset, *J. Microsc.* **1987,** *145*, 61.

Handbook of Microscopy
S. Amelinckx, D. van Dyck, J. van Landuyt, G. van Tendeloo

14 Nuclear Materials

14.1 Introductory Remarks on Handling Nuclear Materials

14.1.1 Overview

Optical and electron optical methods are used in research for characterizing all types of nuclear materials, although the problems treated vary widely. Before irradiation, material properties have to be correlated with the microstructure as the necessary basis from which after in-pile service property changes due to the histories of irradiation, stress, temperature, and corrosion can be established, and improved materials can be developed. The high standards for reliability and safety in the nuclear industry require a detailed characterization of all materials before they are used in nuclear reactors. The methods are different in the analysis of radioactive particles on filters or in biological tissue (see Sec. 14.5 of this Chapter).

Necessary background information on reactor technology and nuclear fuel cycle can be found in *Ullmann's Encyclopedia of Industrial Chemistry* [1]. Structural materials, nuclear fuels and materials for waste disposal, are treated by Frost [2].

From a practical point of view, nuclear materials have to be classified according to their degree of radioactivity. This has to be quantified along three lines:

(i) Out-of-pile by the rate of nuclear decay, measured in becquerels (Bq) ($1\,\mathrm{Bq} = 1\,\mathrm{s}^{-1}$) (formerly expressed in curies (Ci), $1\,\mathrm{Ci} = 3.7 \times 10^{10}\,\mathrm{Bq}$), with the emission of γ quanta, α and β particles, and in-pile by the rate of nuclear fission with the emission of energetic fission fragments, neutrons, and γ quanta.

(ii) By the energies of the various particles absorbed in biological tissue or in solids. The absorbed energy is defined by the gray ($1\,\mathrm{Gy} = 1\,\mathrm{J\,kg}^{-1} = 100\,\mathrm{rad}$).

(iii) Depending on the type of radiation and the type of biological tissue the absorbed energy dose (gray, Gy) has to be multiplied with a factor f to arrive at the biological effects. Formerly these were qualified by the *r*adiation *e*quivalent dose for *m*an (rem). Now the official unit is the sievert (Sv), and $1\,\mathrm{Sv} = 100\,\mathrm{rem}$.

The general rules for the safe handling of radionuclides have been established by the International Atomic Energy Agency (IAEA) in Vienna [3], apart from national law. A useful collection of rules and data on health physics, safety, and control

measures in handling transuranium elements has been given by Vaane [4]. Further details on these and other aspects of handling radioactive materials are treated in the IAEA *Safety Series* [5].

Because of the possible health hazards in work with radioactive materials, as with other toxic or otherwise dangerous materials, special permission is required, and has to be supervised by a qualified authority. Based on local and national laws, this authority usually defines the degree of protection and the quantity and type of radioactive material allowed for handling and storage in a given type of installation.

There are roughly three types of protective installations for radioactive materials, to which the optical and electron optical instrumentation has to be adapted:

(a) Irradiated nuclear fuel elements and the fuel rods can only be remotely dismantled, inspected, cut and analyzed in α/γ cells. These are heavily shielded concrete cells which contain air-tight α boxes equipped for the introduction of manipulators for handling equipment and materials.
(b) Structural alloys (not in previous contact with nuclear fuels) in general contain isotopes which are γ activated by the neutron flux in the reactor core. Hence, ordinary sized specimens of such irradiated structural core materials can be manipulated in simpler γ concrete cells or behind lead shielding provided their surface is not α contaminated.
(c) Materials containing purely α-emitting heavy isotopes, such as ^{235}U and ^{239}Pu, under all circumstances have to be stored in air-tight containers and manipulated in air-tight glove boxes. If the α material contains isotopes emitting hard γ rays, for example certain plutonium, americium, and curium isotopes, the glove box has to be shielded as well.

In general, the procedure to prepare and to analyze radioactive specimens by advanced analytical techniques is as follows. The heavily shielded α/γ or γ concrete cells are used for cutting smaller parts and raw specimens from highly radioactive fuel rods and from irradiated structural materials. Some of this material may be submitted to further testing in these cells. Eventually, depending on specimen size, the final preparation and microscopic analysis can be done under conditions which require less severe or even very little γ shielding.

14.1.2 Technical Solutions for Optical and Electron Microscopes

Here the methods for analyzing highly α/γ-active specimens are mainly outlined. The analysis of irradiated γ-active structural alloys without α contamination is correspondingly simpler. The examination of highly α/γ-radioactive samples by microscopy is complicated by two factors:

(i) Total α-containment is required at every stage of specimen preparation, handling, transfer, and observation to prevent possible α contamination of the environment and of the operator.
(ii) Adequate γ protection of the operator is essential to minimize personal exposure to γ radiation from the samples.

14.1.2.1 Optical Microscopy

For the analysis of the cross-sections of fuel rods and of macroscopic samples from reactor core materials (steels, nickel-base alloys, Zircaloy, etc.), remotely operated optical microscopes are attached to the α/γ and γ hot cells.

For the analysis of nonirradiated, plutonium-containing fuel materials (ceramic pellets or alloys) the optical microscope is enclosed wholly or partly in an air-tight α box. The specimens have to be transferred from the preparation box without breaking the air-tight containments by a special technique (double-lid air locks or welded plastic bags) into the α box of the microscope.

In both of the above situations all conventional techniques of optical microscopy can be applied up to 1000× magnification, as outlined in Secs. 14.2 and 14.3 of this Chapter.

14.1.2.2 Transmission Electron Microscopy

Due to the small size of transmission electron microscopy (TEM) specimens, and with the shielding of the specimen chamber by the microscope column and by the coils of the objective lens, the operator is totally protected against γ radiation from the specimen. Also, the final stage of specimen thinning poses little problems for radiation protection as long as the specimens are neutron-irradiated materials (alloys or ceramics) free of α emitters. These near to normal conditions are, however, drastically changed if irradiated nuclear fuels or nonirradiated alloys and ceramics containing plutonium and higher actinides are to be analyzed by TEM (see Sec. 14.3 of this Chapter).

In the following, a transmission electron microscope for work with highly α/γ-active specimens at the European Institute for Transuranium Elements, Karlsruhe, is described. It is a basically standard 200 keV Hitachi H700 HST with a scanning attachment for scanning transmission and secondary electron imaging and a Tracor Northern TN 5500 Energy Dispersive X-ray Analysis and image-processing system. The modifications made in order to meet the requirements for operation with radioactive and α-contaminated materials are as follows:

(i) Filters (glass fibre, 20 $m^3 h^{-1}$ capacity) were fitted in the rough vacuum lines between the rotary pumps and the microscope to prevent any aerosol release from the vacuum system. This necessitated replacing the original pumps with pumps of a higher pumping speed (1370 $l\,min^{-1}$). The fore vacuum pumps exhaust into a waste gas system which is continually monitored for $\alpha/\beta/\gamma$ radiation levels.

(ii) A self-contained recirculating cooling water circuit was attached to the microscope.

(iii) The side entry goniometer stage was rebuilt with minor internal modifications to make it α tight, but this involved no changes in the design or principle of operation. No additional restriction is placed on the range of specimen movement or tilting.

(iv) The microscope is linked to a three-stage glove box chain for the changing and handling of specimens. This system is shown in Fig. 1. A three-stage chain was chosen so as to minimize the

Figure 1. The Hitachi H700 HST transmission electron microscope linked to a three-stage glove box system for the handling of radioactive and α-contaminated specimens. Samples are introduced into glove box 1 through the transfer device at A, and loaded into the specimen holder at B, where just the tip of the holder projects into the contaminated glove box. Glove box 2 is for transfer and specimen holder storage. Glove box 3 links directly to the microscope through a double door and bellows arrangement.

possible transfer of contamination to the microscope column, since the column must be opened at regular intervals for servicing and cleaning.

The as-prepared specimens are introduced into box 1 through a specimen transfer chamber (A), which is compatible with flanges on the specimen preparation glove boxes. In this glove box they are loaded into the microscope specimen holder under a stereo microscope, which is also equipped with a camera. Only the very tip of the holder projects into box 1 through an airlock arrangement (B), which is the only connection between glove box 1 and the next box, and is closed when no holder is in position. The system is shown in detail in Fig. 2.

Box 2 serves as a transfer box, and also as storage for the various different microscope specimen holders. It is connected by a simple lifting door to box 3.

Glove box 3 can be opened to the electron microscope goniometer stage using a double door and bellows arrangement as shown in Fig. 3. This double door is opened for specimen changing, but

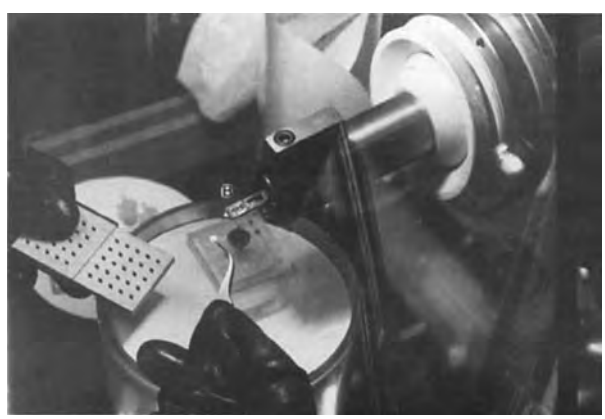


Figure 2. The tip of the specimen holder projects into glove box 1 through a small aperture at position B, which can be sealed with a ball valve when not in use. This system is designed to minimize the transfer of α contamination to the microscope column. Specimens, either on a grid or as 3 mm disk thinned at the center, are loaded with tweezers into the holder under a stereo microscope mounted above the glove box.

Figure 3. The microscope goniometer stage (3) is linked to the glove box system via a La Calhène double door (1) and a bellows (2). This double door is detached, as shown, when the microscope is in use, but attached and the inner component opened for access to the goniometer stage.

closed and decoupled from the microscope column during observation. This decoupling is essential since vibration transfer from the recirculating nitrogen atmosphere of the glove boxes strongly affects the image resolution.

The glove boxes are equipped with automatic alarms for loss of underpressure, fire, and flooding, and with operator hand/foot alarms for accidents. The laboratory air and the glove box nitrogen recirculating system are continuously monitored for contamination.

Operations such as plate changing, filament replacement, or aperture renewal are performed in the normal way, but with careful monitoring at each stage for possible contamination. In over 10 years of daily operation with radioactive materials no significant levels of contamination have ever been detected in the microscope column. The major exposure risk is during the specimen loading and transfer operations. Here it has been found by experience that the total operator dose is smaller when the operations are performed directly, but very rapidly, using gloves and tweezers rather than when they are carried out indirectly, but slowly, using manipulators and a shielded glove box. In the second case there is a much higher risk of specimen damage necessitating repeated preparation work. During specimen loading the operator is protected by wearing a lead jacket, and the radiation dose is monitored with finger dosimeters. A typical specimen change operation takes about 15 min.

In transmission mode the attainable image resolution is limited in the first instance by the nonideal nature of the specimens rather than the microscope performance, but a routine resolution of better than 0.1 nm is obtained, with a demonstrated capability of 0.2 nm. In secondary electron mode a surface resolution of better than 5.0 nm is achieved.

14.1.2.3 Scanning Electron Microscopy

A JEOL JEM 35C Scanning Electron Microscope is being used at the European Institute for Transuranium Elements for the examination of α/γ-active samples, including full cross-sections of irradiated plutonium-containing fuels. In this case the approach to the problems introduced by the active nature of the samples was quite different—the column of the microscope was totally separated from the

electronics and control console, and mounted complete with pumps inside a heavily shielded cell. The system is illustrated in Fig. 4.

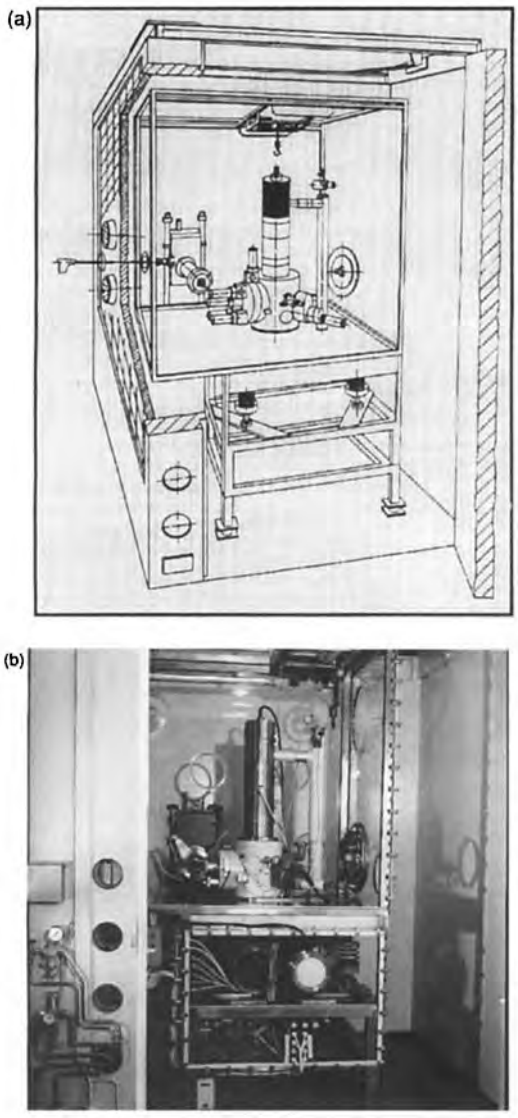


Figure 4. (a) Schematic view of the detached column and pumping system used to contain the JEOL JEM 35C Scanning Electron Microscope inside a glove box and a heavily shielded lead cell with one wall movable. (b) The JEM 35C microscope was installed before the closure of the glove box. The system was then moved into the lead cell [6].

The microscope column and specimen chamber are mounted inside a large fully sealed N_2 atmosphere glove box for α containment, which is surrounded on either side by movable 15 cm thick steel walls and at the front by a 15 cm thick lead brick wall for γ shielding. The front wall contains two windows for visual inspection of the upper part of the box, and manipulators for specimen handling. The steel walls can be swung aside to provide operator access to the glove box for servicing operations. The maximum permissible γ activity of the sample is 2 Ci.

The high-vacuum pumps are mounted inside the cell and are backed through filters by rotary pumps outside. A titanium ion-getter pump is used for the electron gun, and a turbomolecular pump for the specimen chamber.

A shielded secondary electron detector has been installed, and to compensate for the losses in signal involved a high-brightness LaB_6 electron gun has been used.

All the movements of the goniometer stage are driven by micromotors controlled by potentiometric transducers.

Samples are prepared by fractography and are not decontaminated. When necessary they can be coated using a remote-controlled sputtering chamber.

A resolution of 5.0 nm can be routinely achieved, but for large irradiated fuel samples this is reduced by the specimen activity.

For routine applications of scanning electron microscopy (SEM) to γ-active samples free of α contamination a much simpler installation can be used. The microscope column and pumps are freely accessible and only shielding between the specimen chamber and operator is required. Samples must, however, be

decontaminated in a shielded cell before observation to prevent contamination of the microscope from loose active particles.

14.2 Structural Materials (Alloys)

14.2.1 Introduction

The principal microscopic method applied to the characterization and further development of technological alloys for use in nuclear reactors is TEM in combination with metallography, and SEM in combination with mechanical testing for the analysis of fracture surfaces as well as for further surface studies. In most cases the relation between microstructure and the properties of the alloy before service has to be compared with this relation after in-pile service. This means that the methods of investigation must provide quantitative results. Therefore, in Sec. 14.2.2 of this Chapter an outline of several aspects particular to the quantitative analysis of irradiated alloys mainly by SEM and TEM is first given.

Next, in Secs. 14.2.3–14.2.6 of this Chapter, six types of alloys are discussed as they are used in the most common applications in nuclear technology. These are:

- low alloyed steels for reactor pressure vessels (RPVs) and components of the primary coolant circuit;
- ferritic and austenitic steels for piping systems;
- high corrosion-resistant and high-strength nickel-base alloys (e.g., for steam generator tubing);
- zirconium-based alloys for cladding in light-water reactors (LWRs) and hafnium;
- cladding and wrapper materials for fast breeder reactor (FBR) fuel elements; and
- first wall and blanket structures in fusion devices.

For most of these alloys the main characteristics which have to be analyzed by microscopic methods are condensed in the relevant sections in the form of four tables:

(1) composition (specifications),
(2) thermomechanical treatment,
(3) main applications and in-service conditions, and
(4) examination procedures before and after service,

with short descriptions of the relevant structural properties and some details illustrated by micrographs.

14.2.1.1 Microscopy of Irradiated Alloys

The quantitative aspects of the various methods of microscopy are treated in Vols. 1 and 2 of this Handbook. Hence, only some aspects of metallography and examples of quantitative TEM and SEM particular to irradiated alloys in general are given before the six groups of alloys are treated.

Metallography

The metallography of irradiated structural alloys (steels) is based on the wide experience gained with the nonirradiated reference materials. However, a few particulars in the metallography of irradiated alloys

should be mentioned:

- In the hot cells and glove boxes the use of dangerous solutions containing, for example, critical concentrations of perchloric and picric acid is to be avoided for safety reasons.
- When comparing the surface attack by a given etching solution between the irradiated alloy and its nonirradiated reference state the irradiated alloy is often attacked faster. Hence the time of attack should be shorter for the irradiated material to reach the same degree of attack as for the reference material.
- Under neutron irradiation, alloys containing boride precipitates show a special effect due to the nuclear reaction $^{10}B(n, \alpha)^{7}Li$. The high-energy α particle and the heavier lithium recoil atom emitted from the location of the boron atom are slowed down in the adjacent matrix, producing additional lattice damage (Frenkel pairs and clusters). If the irradiation temperature is not too high, that is, below $0.5T_m$, this damage does not anneal out under irradiation and can be made visible by etching. If the borides are mainly precipitated at grain boundaries these become decorated by a double line with the grain boundary in between. The distance of each line from the grain boundary indicates the range of the emitted particles, as shown in Fig. 5.
- Neutron irradiation of the natural isotopes of iron, cobalt, and nickel involves (n, γ) reactions. Amongst these, the decay of the ^{60}Co nuclei is accompanied by especially hard γ radiation (>3 MeV), with a half-life of 5.25 years. Following the different nuclear properties of the iron and nickel natural isotopes, a neutron-irradiated austenitic 16CrNi12 steel gets a three times higher γ dose rate than an equivalent amount of martensitic 10–12% Cr steel after the same neutron dose.

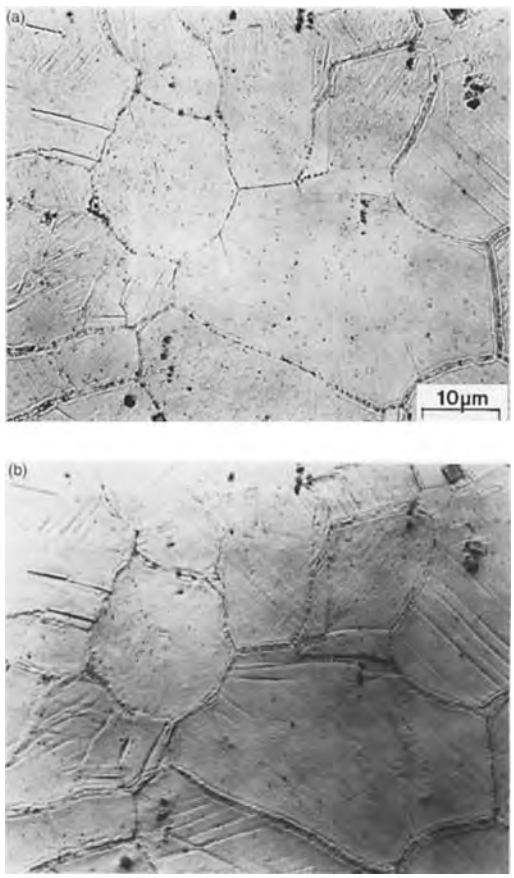


Figure 5. The recoil zones of the $^{10}B(n,\alpha)^{7}Li$ reaction along the grain boundaries in DIN 1.4970 steel (X10 Cr Ni Mo Ti B 15 15)revealed by sensitive etching. Irradiation dose 2 dpa, 100 appm He. (a) Standard optical micrograph. (b) As (a) but with interference microscopy.

In addition to the etching methods mentioned below for the individual alloys and purposes in Tables 1–3, a universal etching solution for irradiated steels is given in Table 4 which can be used for both martensitic and austenitic materials.

Table 1. Chemical composition of the steels [1]

Materials	Specified analysis (%)									Remarks	Standards, regulations
	C (max.)	Si	Mn	P (max.)	S (max.)	Ni	Cr	Co (max.)	Mo		
SA 508 cl. 2	0.27	0.15–0.4	0.5–1.0	0.025	0.025	0.5–1.0	0.25–0.45		0.55–0.7	V ⩽ 0.05	ASME
22 NiMoCr 3 7	0.16–0.21	0.15–0.35	0.5–1.0	0.02	0.015	0.5–0.9	0.2–0.5	0.03	0.45–0.7	Cu ⩽ 0.20 Al 0.01–0.04	VdTÜV-WB 366 (5.72)
SA 533 Gr B	0.25	0.15–0.4	1.15–1.5	0.035	0.04	0.4–0.7			0.45–0.6		ASME
16 MND 5	0.2	0.10–0.30	1.15–1.55	0.015	0.012	0.5–0.8	⩽0.25		0.45–0.5	V ⩽ 0.03 Cu ⩽ 0.20 Al ⩽ 0.04	RCC-M
20 MnMoNi 5 5	0.17–0.23	0.15–0.30	1.2–1.5	0.012	0.008	0.5–0.8	⩽0.20	0.03	0.4–0.55	Cu ⩽ 0.12 V ⩽ 0.02	KTA 3201.1
15 Cr 2 MFA	0.13–0.18	0.17–0.37	0.3–0.6	0.02	0.02	0.4	2.5–3.0		0.6–0.8	V 0.25–0.35 Cu 0.20 ± 0.05	Russian regulations
15 Cr 2 NMFAA	0.13–0.18	0.17–0.37	0.3–0.6	0.01	0.012	1.0–1.5	1.8–2.3		0.5–0.7	V 0.1–0.12 Cu ⩽ 0.10	Russian regulations
15 MnNi 6 3	0.12–0.18	0.15–0.35	1.2–1.65	0.015	0.005	0.5–0.85	⩽0.15		⩽0.05	V ⩽ 0.02, Ti ⩽ 0.02 N ⩽ 0.015	VdTÜV-WB 427
SA 516 Gr. 70	0.28	0.15–0.4	0.85–1.2	0.035	0.04						ASME

ASME, American Society of Mechanical Engineers, VdTÜv-WB, Vereinigung der Technischen Überwachungsvereine-Werkstoffblatt, Part 'Mechanical Components of PWR nuclear; RCC-M, Règles de Conception et de Construction.
KTA, Kerntechnischer Ausschuss.

Quantitative Fractography by Scanning Electron Microscopy

Irradiation embrittlement of reactor materials is an important phenomenon. It can be analyzed by comparing the mechanical properties and microstructure of the irradiated alloy with reference material submitted to the same thermal history as the irradiated alloy. Mechanical testing of the irradiated material is performed in remotely handled hot cells. Following

Table 2. Materials specification and heat treatment sequences

	NiMoCr type	MnMoNi type	Cr, CrMo type
Material specifications			
KTA regulations [20]	3201.1, app. 1	3201.1, app. 1	Russian regulations (originally Soviet)
ASME [21]	SA508, class 2	SA533, grade B, class 1 SA508, class 3	DIN 17243, ASME 182 F22
Heat treatments			
(typical range for steel)	22 NiMoCr 3 7	20 MnMoNi 5 5	10 CrMo 9 10
Quenching	880–920°C/oil, air	870–930°C/water	900–960°C/oil, air
Tempering	620–700°C	630–680°C	650–750°C
Stress relief			
Annealing (welds)	550–620°C	580–620°C	650–750°C
Soaking			
(hydrogen effusion)	>200°C	260–300°C	

Table 3. Applications, service conditions, problem areas, and targets of characterizations

	NiMoCr type	MnMoNi type	Cr, CrMo type
Reactor type	PWR, PHWR, BWR	PWR, PHWR, BWR	VVER (PWR)
Components	Pressure containing wall of the primary circuit, e.g. reactor pressure vessel, steam generator, main coolant pump casing, pipes, and connections to the auxiliary systems		
Material properties of advance [22, 24]	High strength/ductility at ambient and elevated temperatures, uniform properties across the wall even at high wall thicknesses (up to 600 mm!), excellent weldability with uniform properties in the base metal, heat affected zone (HAZ), and weld metal, low sensitivity to changes in the material properties in the base material, the HAZ, and the weld during service life		
Operational conditions			
Operational temperatures (K)	$\leqslant$623 (543–593)		
Coolant	PWR: light water (containing about <2.5 ppm Li as LiOH, $\leqslant$ 1000 ppm of B as boric acid, $O_2 < 0.5$ pbb, $H_2 \approx 30\,cm^3\,l^{-1}$ BWR: light water, $O_2 \approx 200$ ppb or lower, depending on the special water chemistry treatment (e.g., with hydrogen water chemistry) PHWR: heavy water (D_2O) ($O_2 < \approx 2$ ppb, depending on the special water chemistry, hydrogen injection) The ferritic steel material is separated from the water entirely by a corrosion-resistant cladding		
Coolant pressure (bar)	PWR $\leqslant$ 158	BWR $\leqslant$ 88	PHWR $\leqslant$ 115
Maximum neutron fluence	$\approx 1 \times 10^{19}$ to $6 \times 10^{19}\,cm^{-2}$ depending on the reactor type and on the chemical composition of the steel [26]		

BWR, boiling-water reactor; PHWR, pressurized heavy water reactor; PWR, pressurized water reactor; VVER, water-

Table 4. A universal etching method for irradiated steels [7][a]

Alloy (irradiated steels)	Time	Purpose
1. Martensitic, 10–12% Cr	40 s	Contrasting structural constituents
2. Ferritic, 2.5% Cr	4 min	
3. Austenitic, 16CrNi12	160 s	
4. Welding seams between martensitic and austenitic steels	80–160 s	

[a] Solution: 10 g of $AlCl_3$, 30 ml of HCl, 15 ml of HNO_3, 100 ml of H_2O.

tensile testing, notch impact tests, or creep tests and macroscopic evaluation of the specimen geometry, fractured parts are decontaminated and fixed on the SEM specimen holder. Loose particles are removed by a brief ultrasonic cleaning to avoid the loss of precipitates located in the fracture surface. According to the type of load, fracture is classified as follows:

(1) Fracture after mechanical load:
  (a) spontaneous fracture (ductile or brittle);
  (b) fatigue fracture.
(2) Corrosion-induced cracks and fracture.
(3) Thermally induced cracks and fracture.
(4) As (3) and irradiation induced cracks and fracture.

Fracture cannot be observed directly, hence the process of material separation has to be deduced from the fracture pattern. It is of crucial importance to recognize fracture initiators as well as to determine the deformability of the matrix or the residual fracture. Each type of fracture is characterized by a number of features which allow conclusions with regard to the load, the medium, and the state of the material before fracturing. These are:

(a) mechanical data (e.g., yield stress, tensile strength, fracture stress and strain, neck diameter);
(b) ductile/brittle proportions of the fracture surface (SEM);
(c) fracture roughness observed from metallography of a cut section or by confocal laser scanning microscopy of the fracture surface;
(d) dimple structure and statistics by ECD (equivalent circle diameter);
(e) precipitates in the dimples (SEM plus EDX or WDX); and
(f) dislocation structure and precipitate distribution in the starting material (TEM plus EDX).

Basically, the effects of neutron irradiation are expected to possibly produce:

(1) changes of the dislocation distribution, for example due to the effects of Frenkel defects, defct agglomerates and subgrain formation;
(2) changes in precipitate sizes, compositions, and density; and
(3) formation of helium bubbles and/or voids.

To recognize and characterize irradiation-induced effects, a quantitative evaluation of the fracture surface has to be carried out by comparing a suitable selection of items (a) to (f) between the irradiated material and the reference material.

In this context the applications of EDX and WDX to irradiated alloys meet certain limitations. Due to the neutron activation

of certain alloy constituents, especially nickel and iron, the γ activity of the specimen due to the spontaneous radioactive decay interferes with the excitation of the characteristic X rays by the electron beam of the scanning electron microscope. In the small Si(Li) counting crystal the hard γ radiation generates a number of interfering pulses, resulting in an excessive counting rate and high dead times. Hence the background is increased considerably and the measuring time accordingly. In an unirradiated specimen of the low-alloyed constructional steel 22 NiMoCr 3 7 (DIN 1.6751) the nickel content of about 1% (detection limit $\geqslant 0.1\%$) is clearly visible in the spectrum after a measuring time of 1 min. In a slightly irradiated specimen (fluence ~ 0.5 mSv) a measuring time of at least 5 min is required to detect a nickel peak above the background.

Other interferences are caused by nuclides such as ^{55}Fe, which emits a γ quantum of 5.95 keV located exactly on the manganese peak. This means the manganese content for iron alloys cannot be determined unambiguously. Of course this spectrum is also generated without the exciting SEM electron beam. Therefore, this low-energy γ radiation can be used to calibrate the EDX detector, and the resolution of the manganese peak generated by an ^{55}Fe test emitter can be checked.

In highly activated materials K capture and β radiation contribute to the background, and EDX is no longer suitable. In this case WDX is more useful as the analytical radiation is deflected by the spectrometer prism and, hence, the detector does not directly see the irradiated specimen. Thus, a better peak to background ratio is achieved, and the resolution is improved. However, the spectra are recorded in wavelength steps, and several crystals have to be used to cover the necessary range of wavelengths; the method is thus time-consuming.

Most of the literature on quantitative fracture analysis [8–12] deals with the evaluation of the fracture roughness or dimple structures in ductile fracture. To characterize the change of the material ductility it is useful to compare the dimple size distribution. In this case the form of an ECD can be applied, and varying dimple diameters can be evaluated statistically. If inclusions or precipitates still exist inside the dimples and can be analyzed by means of WDX or EDX, information can be deduced with respect to cavity initiation. Fracture initiation may be changed by neutron irradiation if the microstructure and its precipitation behavior are affected.

As far as martensitic 12% Cr steel is concerned, this phenomenon has been observed after in-pile creep. Neutron irradiation was found to result in a shorter time to fracture and a hardening of the alloy as compared to reference creep tests without irradiation [13].

On the basis of the statistics of the dimple structure and the analysis of the inclusions and precipitates, it is concluded that in-pile conditions mainly cause the formation of large dimples. These are the dimples generated around inclusions (MnS, Al_2O_3) or primary carbides (NbC). The other precipitates ($M_{23}C_6$) along the martensite laths were coarser, but their influence on dimple formation in fracture was much smaller than in the unirradiated alloy.

The effect on the ductility of the alloy is derived from the dimple depth. This can be obtained from metallographic

cross-sections. Another possibility of evaluating surface structures or roughness consists in the application of confocal laser scanning microscopy. By recording level lines a three-dimensional image of the surface can be set up and evaluated quantitatively (xy resolution 0.2 μm and z resolution 0.4 μm) [14].

Martensitic 10–12% Cr steel (1.4914) irradiated at $T \geqslant 400°C$ shows ductile fracture up to high neutron doses $\geqslant 65$ dpa. Compared to the unirradiated alloy, irradiation decreased the upper shelf of impact tests and the surface area of the ductile fracture showed the reduced dimple depth (Fig. 6). The coarsening of

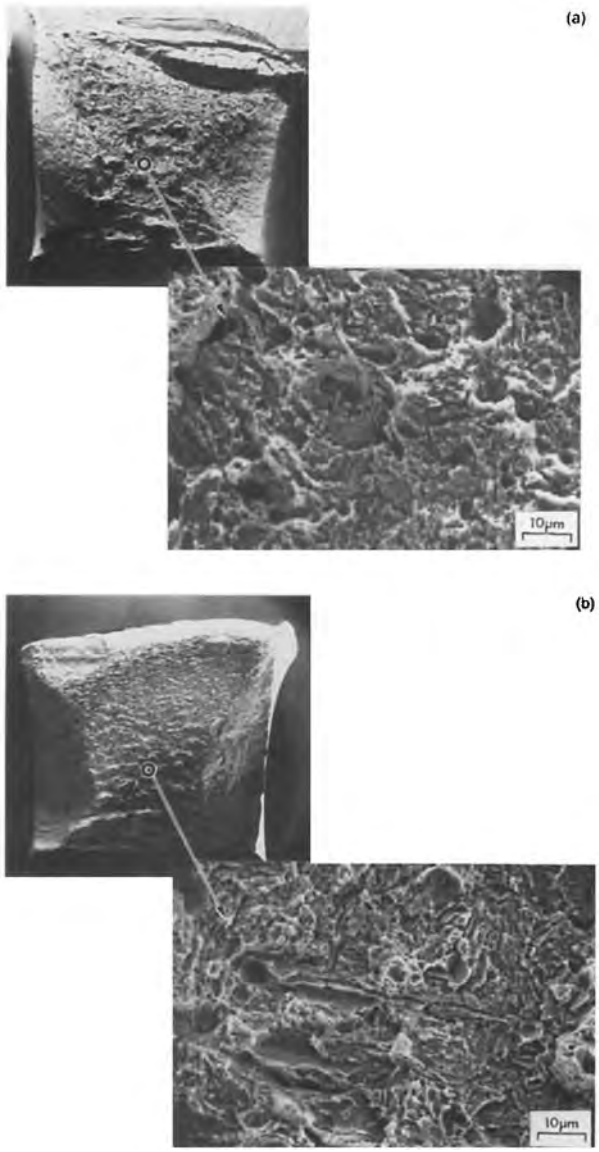


Figure 6. Notch bar tests of martensitic 10–12% Cr steel (DIN 1.4914). (a) Ductile fracture of an unirradiated sample tested at −30°C. (b) Ductile fracture of an irradiated sample tested at +60°C.

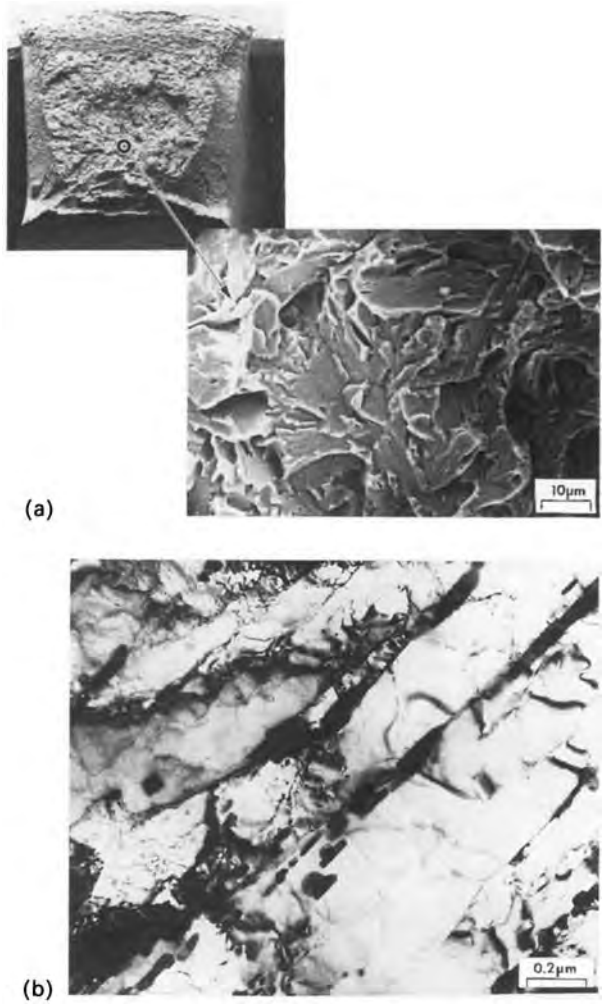


Figure 7. Cleavage fracture of a nonirradiated sample (same material as Fig. 6) with $FATT_{50}$. (a) SEM image of the fracture surface after testing at $T = -60°C$. (b) Microstructure of the martensitic 10–12% Cr steel before the test as produced by quenching and annealing for 30 min at 1075°C and subsequently for 2 h at 700°C.

the precipitates along the lath boundaries after irradiation caused a flattening of this fracture surface. Depending on irradiation, the martensitic cleavage changes too. Figure 7 shows the fracture surface and the microstructure of the lath in the non-irradiated state. After irradiation (Fig. 8), a further separation occurred along the lath boundaries, but the crack runs transcrystallinely through the material (Fig. 8a). The matrix is strengthened by coherent, finely dispersed, irradiation-induced precipitates (Fig. 9). The DBTT (ductile to brittle transition temperature) is shifted from −60°C for the nonirradiated material to +20°C with the irradiated one [15].

Quantitative Transmission Electron Microscopy of Fast Reactor Alloys

Fast reactor materials show some typical forms of lattice defects. The rate of displacement of lattice atoms due to the collisions with neutrons of a high flux of

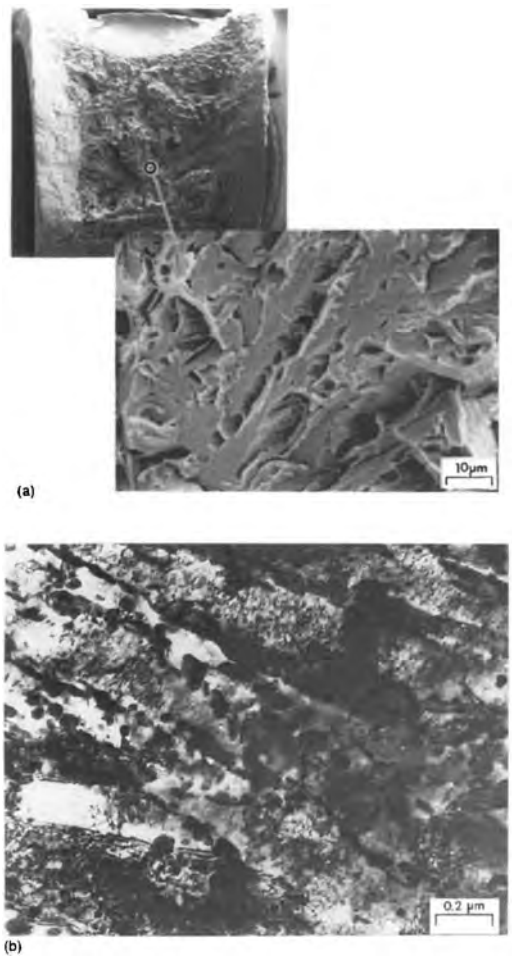


Figure 8. Cleavage fracture of a martensitic 10–12% Cr steel sample (same material as in Fig. 7) after irradiation to a neutron dose corresponding to 74.5–100 dpa. (a) SEM image of the fracture surface with $FATT_{50}$ at a test temperature of +20°C. (b) Martensitic microstructure of the specimen in (a). Note the dense rows of carbides and the presence of several voids.

the order of $10^{15}\,cm^{-2}\,s^{-1}$ results in a high production rate of Frenkel pairs. Some 0.1% of these defects form stable aggregates, vacancies condense into voids, interstitials precipitate in dislocation loops, and second-phase particles may dissolve or new precipitate not normally observed may be generated. By nuclear reactions with the alloy constituents, helium and hydrogen can be formed under certain conditions. The concentrations of these defects have to be controlled because they strongly affect the dimensional stability of the reactor core components and their mechanical properties.

Helium Bubbles. In steels containing boride precipitates the $^{10}B(n,\alpha)^{7}Li$ reaction leads to helium concentrations in the matrix and at grain boundaries from which bubbles are precipitated. Figure 10 shows bubble formation by this effect in the matrix and in a grain boundary of the austenitic steel DIN 1.4970 (Fe–15Cr–15Ni–Ti).

Void Swelling. Void swelling is quantified by measuring the void size distribution, that is, the void concentrations in the various size classes. The visibility of the voids, typically with diameters between 5 and 150 nm, is best in a lattice orientation for kinematic diffraction conditions when large diffraction contrast is absent. A slight underfocus gives the facetted voids a dark rim, and this improves the visibility. A size class is best determined by comparison with a tuneable equisized diaphragm coupled to a multi-class counter. At small void sizes it is important to remain in focus as near as possible to observe the correct size. The automatic measurement by image-processing methods mostly fails due to the strongly varying contrast conditions.

Voids may be nucleated at existing defect clusters (helium atoms and nuclei of second-phase particles). Figure 11 shows facetted voids, part of which have grown together with NbC precipitates.

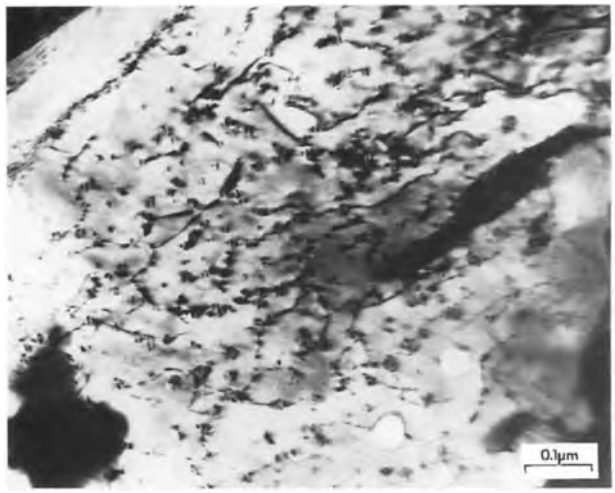


Figure 9. TEM image of the matrix of the sample of Fig. 8. The irradiation has hardened the matrix by producing fine and coherent precipitates.

For the determination of the volume concentration of the voids the foil thickness has to be measured. Here different methods have to be used, especially at the beginning of the study of a new material as long as the operator is not acquainted with the typical aspects of this material:

(i) analysis of stereo pairs of features localized on the upper and lower specimen surface such as dislocation ends or dirt spots;
(ii) the extinction fringe method;
(iii) the use of a contamination cone produced by a steady highly focused electron beam in a 45° tilt position and measuring the distance between the upper and lower parts in the −45° position; and
(iv) by convergent beam diffractometry [16, 17].

Yet, contamination methods usually lead to somewhat larger thickness values

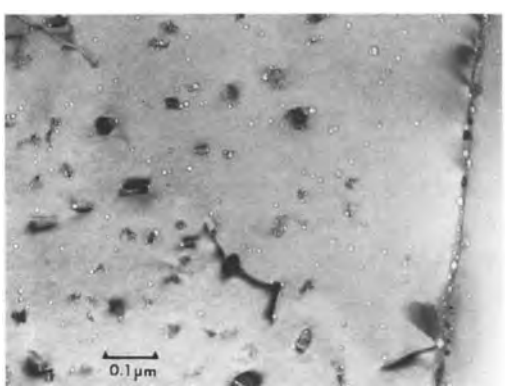


Figure 10. Helium bubbles formed at and in the vicinity of grain boundaries due to nearby boride precipitates in the austinitis steel DIN 1.4970 irradiated at 640°C to a dose of 5 dpa.

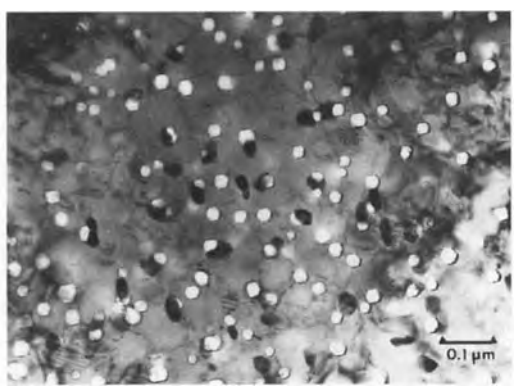


Figure 11. Co-precipitation of voids and NbC in a DIN 1.4981 steel irradiated at 465°C to a dose of 65 dpa.

of the foil than diffraction contrast-controlled methods. All these methods are quite laborious and time-consuming, and an experienced operator is able to estimate the thickness of a specimen of a well-known material to a degree of accuracy which is nearly comparable to the methods mentioned above. In addition, comparison of the TEM analysis with the macroscopically determined density of the material provides a control of the measurements.

Dislocation Loops. In contrast to the void analysis, the determination of the nature and volume concentrations of dislocation loops is more difficult. The visibility of the loops is best under two beam diffraction conditions with $\boldsymbol{gb} \neq 0$. For the distinction between interstitial and vacancy loops the loop size effect caused by the change from inside to outside contrast is used when operating under $+\boldsymbol{g}$ and $-\boldsymbol{g}$ conditions. The size of the loop image is larger (outside contrast) if the value of the product $(\boldsymbol{gb})\boldsymbol{s}$ is positive. Interstitial and vacancy loops have different signs of $\boldsymbol{b}$ according to the FS/RH convention [18].

Figure 12 gives an example of possible interactions between voids, dislocations, and dislocation loops. The loops show in part stacking fault contrast.

Irradiation-Induced Precipitates. On the one hand, the displacements of atoms by collision with fast neutrons and with knock-on atoms on the average increase the mobility of the lattice atoms. Thus, thermodynamic and chemical equilibria are attained in a shorter time. On the other hand, the strongly increased concentrations of lattice vacancies and interstitials give rise to precipitation mechanisms not observed without the strong dynamic irradiation effects. Hence, both the dissolution of precipitates and the formation of new precipitates can be observed.

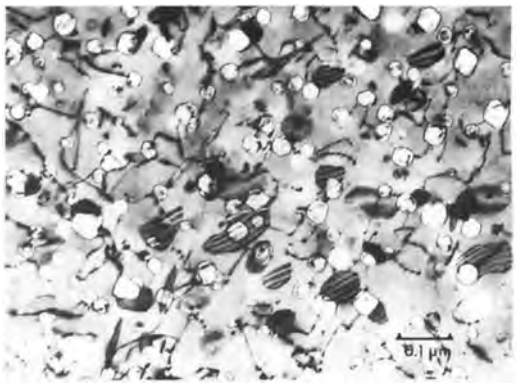


Figure 12. Formation of voids and loops in a titanium-stabilized steel (DIN 1.4970) (cladding material), irradiated at 465°C to a dose of 30 dpa.

An example of a precipitate which is only formed under irradiation is given in Fig. 13 [19]. In an austenitic titanium-stabilized steel (DIN 1.4970), irradiated under a neutron flux of $2 \times 10^{15}\,\mathrm{cm^{-2}\,s^{-1}}$ to a fluence of $48\,\mathrm{dpa_{NRT}}$, the completely coherent γ'-phase Ni_3Si is barely visible under bright field conditions (Fig. 13a). The diffraction pattern (Fig. 13b) shows weak reflections between the normal face-centered cubic (f.c.c.) matrix spots, indicating a phase with the same lattice parameter as the matrix but of ordered primitive Bravais type instead of the f.c.c. matrix. The dark field image (Fig. 13c) obtained by using one of the superstructure reflections readily shows the faint images of the precipitates [19]. In general, these precipitates are found only in nickel-rich steels.

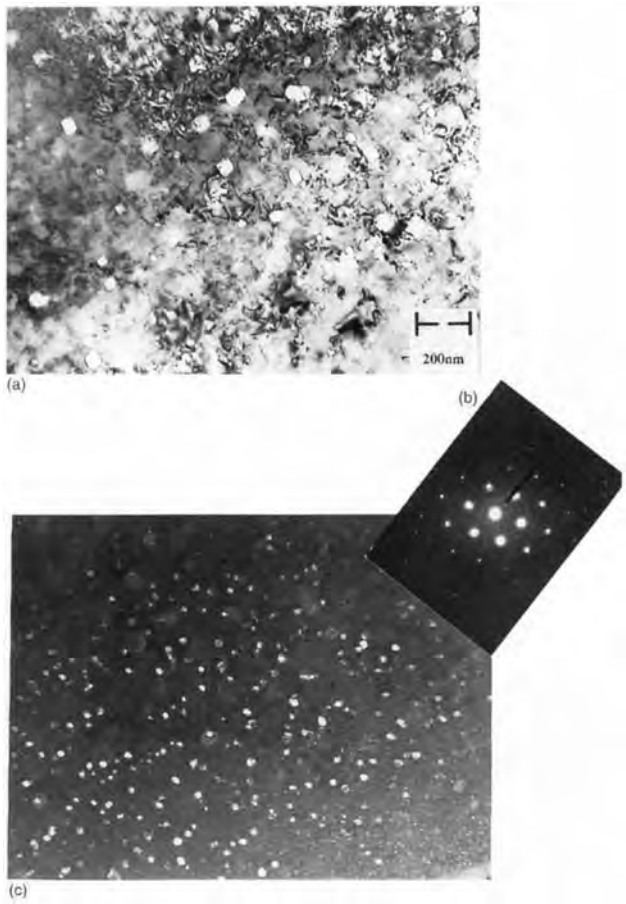


Figure 13. γ' precipitates of type Ni_3Si in a titanium-stabilized steel (DIN 1.4970) after irradiation with fast neutrons at 480°C. (a) The bright field image shows only faint evidence of the precipitates. (b) Diffraction pattern. (c) The dark field image with the superstructure reflection marked in (b) reveals the high concentration of these precipitates.

14.2.2 Reactor Pressure Vessels and Other Main Components of the Primary Circuit

14.2.2.1 Fine-Grained Structural Steels

In general, two different material concepts are employed for the components of the pressurized wall of the primary circuit of LWRs:

(a) All components of the primary circuit (RPV, main coolant line, pump, steam generator) are made of low-alloy ferritic steels. Component surfaces in contact with the primary coolant have a weld-deposited or a rolled cladding.

(b) Low-alloy ferritic steels are used for the large components (RPV, steam generator, pressurizer) in addition to their use for corrosion-resistant cladding. Smaller components, for example the main coolant piping, main coolant pumps, and gate valves, are made of austenitic stainless steel as the structural material.

This section will describe the fine-grained low-alloy ferritic steels used in both concepts [20–25]. The austenitic stainless materials will be described in Sec. 14.2.3.

The steels used are mostly of the nickel–molybdenum–chromium type (22 NiMoCr 3 7) or the manganese–molybdenum–nickel type (20 MnMoNi 5 5), in the USA and Europe, and the chromium–molybdenum type in Russian (originally Soviet) designs.

All steels exhibit a high fracture toughness over the entire service temperature range. The nominal chemical composition of the most important materials is given in Table 1 (for more detailed information the reader is referred to *Ullman's Encyclopedia of Industrial Chemistry* [1]). The specifications and heat treatment procedures of the materials are described in Table 2. The applications, service conditions, problem areas, and targets of characterizations are surveyed in Tables 3 and 5.

14.2.2.2 Characterization and Evaluation of the As-Fabricated State

As can be deduced from Table 5, the purpose of a microscopic examination is focused on various targets, as described below.

Microstructural Properties

Specified values of microstructural properties, for example the size and number density of nonmetallic inclusions, segregations, grain size, and hardness, are checked.

The impurity characteristics, among others, affect the ductility, the fracture toughness, and the anisotropic behavior of the material. In addition, the inclusions, which are elongated according to the material flow during deformation, are often accompanied by microsegregations arranged along the striations, and may cause problems after welding or service due to delayed hydrogen cracking after welding [28, 29] or to lamellar tearing, where cracking occurs in steps along the lamellas enriched with inclusions with lower ductility. To avoid problems, the amount of impurities is restricted. In addition, the risk of lamellar tearing has been reduced by modifications of the design of welds.

Grain Structure

An example of the quenched and tempered base material of 20 MnMoNi 5 5 is shown in Fig. 14. The bainitic structure reflects the quenching and tempering treatment (time at temperature, and cooling rate), and is of important influence with respect to the mechanical properties of the material.

Heat-Affected Zone

The quenched and tempered microstructure of the base material is influenced by heat due to welding. Depending on the distance to the fusion line and the parameters applied during welding, the microstructure of a heat-affected zone (HAZ) can be divided into separate characteristic zones (Fig. 15), starting from the unaffected base material, to the fine-grained zone, and to the coarse-grained zone adjacent to the fusion line and the weld metal. The coarse-grained zone itself is subdivided into an alternating sequence of zones, which are affected by the thermal heat influence of following weld layers, frequently characterized by carbide decorations at the austenite grain boundaries (Troostit seams) and zones which are only superheated by one individual weld pass

Table 5. Survey of procedures for the examination and sample preparation of low-alloyed ferritic steels.

Target	Light microscopy	Scanning electron microscopy	Transmission electron microscopy
As-fabricated condition	Impurities (sulfides, oxides) Cracks/flaws formed during fabrication Grain size Grain structure HAZ characterization, segregations, hardness Stress relief cracks in HAZ (tangential sections!)	Appearance of ductile (dimples) and brittle fracture portions (trans- or intergranular) Irregularities (e.g., flaws) or inclusions of impurities) Structure and composition of contaminants on the surface	Details of the microstructure with respect to dislocation arrangement and density Crystal structure, chemical composition of the matrix (microsegregations, concentration profiles) Morphology and chemical composition of carbides (precipitates)
During/after operation	Verification of nondestructive testing indications	Appearance of the ductile and brittle fracture proportions of irradiation test samples (increasing to amount of inter-granular fracture) Characterization of cracks caused by the operation	Temperature- or irradiation-induced changes in the microstructure with respect to the formation of precipitates/segregates, the carbide structure, or dislocation structures Formation of dislocation networks caused by low cycle fatigue
Sample preparation	Standard metallographic grinding and polishing procedure Etchant for microsections: 3 vol.% HNO_3 in ethanol Etchant for macrosections: 10 vol.% HNO_3 in ethanol	SEM: no special procedures needed On fracture surfaces or surfaces covered with oxides the scales may be removed by the endox 214 method [27]: Cathodic polarization of the surface of interest in an aqueous solution of about 5 wt.% HCN (!), 50 wt.% NaOH and organics) $J \sim 200\,\mathrm{mA\,cm^{-2}}$	*Matrix* Standard preparation of TEM thin foil samples by: – Mechanical cutting of about 0.3 mm sections with low deformation rate by diamond or abrasive (SiC) cutting tools – Mechanical thinning by smooth stepwise (2 µm/step) abrasive grinding down to 0.1 mm thickness – Punching of standard disks with 2–3 mm diameter – Jet stream polishing until an about 50–100 µm diameter hole appears in the center of the disk *Polishing agents* I: 90 vol.% acetic acid 10 vol.% perchloric acid Temperature: 12–15°C, $U = 20$–$30\,\mathrm{V}$ II: 90 vol.% methanol 10 vol.% perchloric acid Temperature: −30°C, $U = 30$–$40\,\mathrm{V}$

			Precipitates Standard metallographic preparation Deep etching about 1 min with 98 vol.% methanol, 1.8 vol.% HNO_3, 0.2 vol.% HCl, 40 g picric acid, extraction of the precipitates with a standard replica technique
Optical instruments used	Standard optical microscopes for metallography, air or immersion oil optics, microhardness tester HV0.1 to HV10	Standard SEM with energy dispersive and/or wavelength dispersive (light elements) electron beam analysis system	Analytical TEM with energy dispersive and/or electron energy loss analyzer system, scanning transmission mode (STEM) for X-ray mapping and electron beam analysis system
Magnification range	10:1–1000:1	20:1–5000:1	30 000:1–1 000 000:1

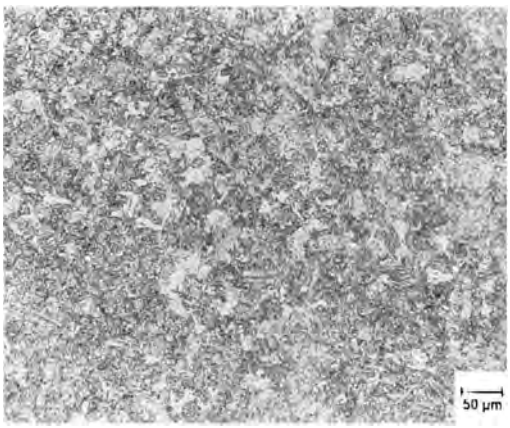


Figure 14. Quenched and tempered microstructure of the steel 20 MnMoNi 5 5.

(e.g. final pass), and therefore reveal a microstructure with a clearly visible primary austenite grain structure formed during superheating, subdivided by lamellar secondary ferrite grains formed during cooling. The uniformity of properties across the weld is checked by hardness profiles (generally Vickers) in macro- (e.g., HV10) or microsections (e.g., HV0.2), and care is taken not to exceed certain limiting values in order to avoid zones with lower ductility and fracture toughness.

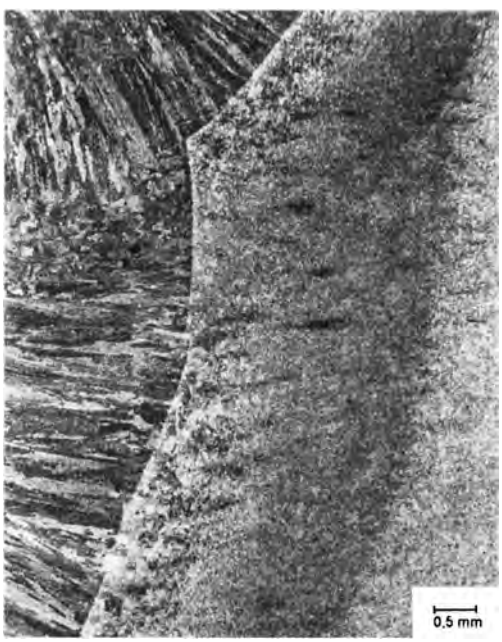


Figure 15. Heat-affected zone of a submerged arc weld of the steel 22 NiMoCr 3 7.

Zones of the HAZ or the weld metal of ferritic steels with high hardness or high residual tensile stresses may also force hydrogen migration. The recombination of atomic hydrogen to the molecular form may cause the formation of hydrogen-induced cracks after quenching or welding (cold cracking). In the HAZ the cracks appear intergranular with respect to the austenite grain boundaries. In the fine-grained base metal, the orientation is often perpendicular to the direction of welding. To reduce the risk of cracking, restrictions on the ingress of hydrogen (drying of electrodes, powders for submerged are welding), the formation of sensitive zones (restricted hardness and grain size), and hydrogen effusion by soaking heat treatments are applied.

During welding, the carbides in the HAZ are dissolved by superheating and precipitated again during cooling. Therefore, an analysis of the morphology and the chemical composition of the carbides, using the extraction replica technique in addition to analytical TEM, not only provides information on the chemical composition of the material, but also on the local thermal treatment of the material (Fig. 16) [30].

The coarse-grained zone of the HAZ is also of importance with respect to the formation of stress relief cracks (SRCs) during the stress relief annealing heat treatment of the weld (Fig. 17). The cracks

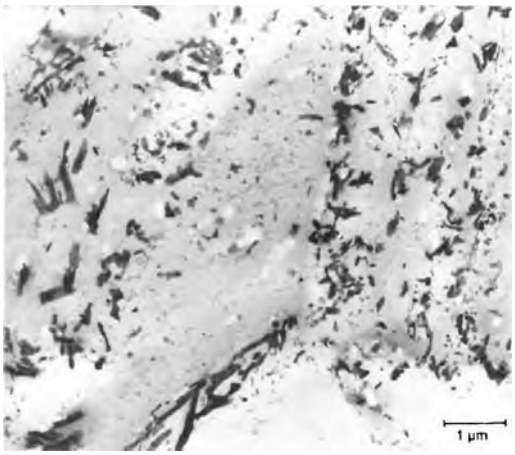


Figure 16. TEM extraction replica of carbides in the coarse-grained zone of an HAZ in 22 NiMoCr 3 7 after stress relief heat treatment. The coarse carbides correspond to M_3C (M = Fe, Mn, Cr), and the fine needles to MC (M = Fe, Mo, Cr, Mn, etc.).

occur intergranularly at the grain boundaries of the austenite grains and are caused by the superposition of creep strain (decrease in dislocation density, relief of residual stresses) with precipitation (carbides) and segregation (sulfur, phosphorus, residual elements) processes, and the existence of coarse austenite grains, which in addition restrict the local creep ductility [23, 29].

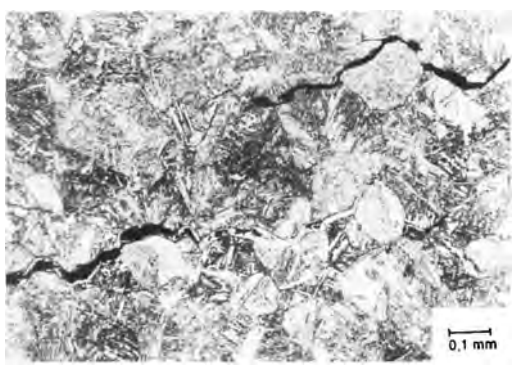


Figure 17. Stress relief cracks in the coarse-grained zone in an HAZ of 22 NiMoCr 3 7 (field metallographic replica technique).

The steel 20 MnMoNi 5 5 is less susceptible to SRCs than 22 NiMoCr 3 7. In both steels, the risk of SRCs is drastically lowered by restrictions on the chemical composition and the welding procedure (heat input, preheat, interpass temperatures) in order to avoid the formation of coarse austenite grains.

14.2.2.3 Characterization and Evaluation During and After Service

Corrosion

As the components are entirely covered by an austenitic stainless steel cladding the materials in general are not affected by corrosion during service. Specific corrosion phenomena of ferritic and austenitic steels are described in more detail in Sec. 14.2.3 of this Chapter.

Low Cycle Fatigue

Many components are operating under conditions resulting in restrictions on the operating lifetime (e.g., thermal stratifications during restricted numbers of start-up or shut-down cycles) caused by low cycle fatigue.

The fatigue straining causes the formation of characteristic dislocation structures (Fig. 18). Their quantitative characterization, for example by TEM, allows the estimation of the residual lifetime of a component [31].

Irradiation-Induced Changes in Properties

During operation under neutron irradiation (and long times at elevated temperature), the material properties may be affected by a shift of the transition temperature regime of the brittle to ductile

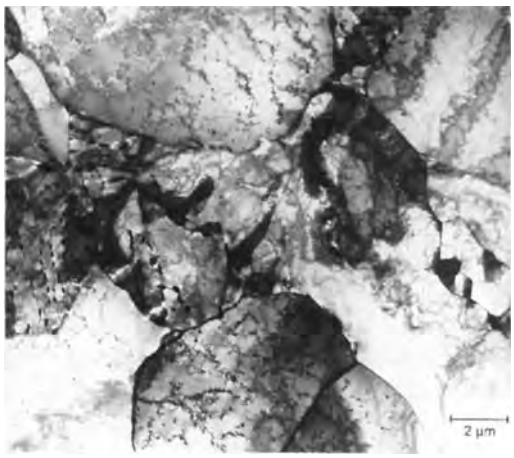


Figure 18. Onset of the formation of dislocation networks at the beginning of low cycle fatigue straining of 20 MnMoNi 5 5 steel (TEM image of a thin foil sample).

fracture behavior to higher temperatures and a decrease of the ductile fracture toughness [26, 32, 33]. As this ageing effect is of importance to the integrity of components, the effect of irradiation is monitored intensively by irradiation programmes for postirradiation mechanical and fracture-mechanical testing.

Recent results have shown that the ageing of the steels caused by neutron irradiation is accompanied by hardening due to the formation of very small and numerous precipitates with diameters between 1 and 10 nm enriched in copper, vanadium, manganese, nickel, and other elements [34, 35]. Because of the characteristics of these precipitates, their examination requires the use of high-resolution TEM and scanning TEM (STEM) techniques, in order to gain additional information on the state of ageing of reactor components.

Thermally Induced Changes in Properties

As most of the components of the primary system operate for a very long time at elevated temperatures, changes in material behavior may be caused by thermal ageing, due to age-hardening processes (nitrogen migration, precipitation, and segregation processes).

The thermal changes in the materials are minimized by the use of deoxidized steels with a low content of degrading impurity elements and adapted tempering and stress relief annealing heat treatments. Laboratory ageing studies [36] and material testing of replaced components after long-term service have shown the low sensitivity of the steels to thermally induced loss of ductility or a shift in transition temperature without affecting component integrity.

14.2.3 Piping

14.2.3.1 Ferritic and Austenitic Steels for Piping Systems

Several ferritic or austenitic stainless steels are in use as materials for pipes, fittings, valves, and pump casings in nuclear and non-nuclear systems (Table 6). For austenitic stainless steel in LWRs, there are generally two types, the unstabilized stainless steels and the stabilized stainless steels. Details on the specifications of commonly used materials and the heat treatment conditions are listed in Table 7. The impact of the use of the individual steels depends on the type of reactor and the system of application (Table 8).

14.2.3.2 Characterization and Evaluation of the As-Fabricated State

The aims of microscopic examination of the different steels, as listed in Table 8,

Table 6. Chemical composition of alloys (wt.%)

Material	C	Si	Mn	P	S	Al	N	Cu	Cr	Mo	Nb	Ni	Ti	V
Ferritic steels[a]														
St 35.8 (1.0305)	≤0.17	0.10–0.35	0.40–0.80	≤0.040	≤0.040									
15 Mo 3 (1.5415)	0.12–0.20	0.10–0.35	0.40–0.80	≤0.035	≤0.035					0.25–0.35				
15 MnNi 63 (1.6210)	0.12–0.18	0.15–0.35	1.20–1.65	≤0.015	≤0.005	0.02–0.055	≤0.015	≤0.06	≤0.15	≤0.05	≤0.004	0.50–0.85	≤0.020	≤0.020
15 NiCuMoNb 5 (1.6368)	0.10–0.17	0.25–0.50	0.80–1.20	≤0.016	≤0.005	≤0.050	≤0.020	0.50–0.80	≤0.30	0.25–0.40	0.015–0.025	1.00–1.30		≤0.020
17 MnMoV 6 4 (1.8817)	≤0.19	0.20–0.50	1.40–1.70	≤0.030	≤0.025		≤0.020			0.20–0.50		0.50–1.00		0.10–0.19
Austenitic stainless steels[a,b]														
X5 CrNi 18 10 (1.4301)	≤0.07	≤1.0	≤2.0	≤0.045	≤0.030				17.0–19.0			8.5–10.5		
X6 CrNiTi 18 10 (1.4541)	≤0.08	≤1.0	≤2.0	≤0.045	≤0.030				17.0–19.0			9.0–12.0	≤0.80 >5% C	
X6 CrNiMoTi 17 12 2 (1.4571)	≤0.08	≤1.0	≤2.0	≤0.045	≤0.030				16.5–18.5	2.0–2.5		10.5–13.5	≤0.80 >10% C	
X6 CrNiNb 18 10 (1.4550)	≤0.08	≤1.0	≤2.0	≤0.045	≤0.030				17.0–19.0		≤1.00 >10%C	9.0–12.0		
X6 CrNiMoNb 17 12 2 (1.4580)	≤0.08	≤1.0	≤2.0	≤0.045	≤0.030				16.5–18.5	2.0–2.5	≤1.00 >10%C	10.5–13.5		

[a] Chemical composition according to DIN specifications.

[b] For applications in nuclear systems, special restrictions on the chemical composition may be specified (e.g., on C (≤0.04%), P, S, Nb, Ti, and Co).

Table 7. Material specification and heat treatment sequence

Material	Specifications	Heat treatment condition	Post-weld heat treatment (°C)
St 35.8	DIN 17175 (1.0305)	Normalized, 900–930°C	520–600
15 Mo 3	DIN 17175 (1.5415)	Normalized, 910–940°C	530–620
15 MnNi 63	(1.6210) KTA 3211.1, app. A1	Normalized, 880–960°C	530–580
15 NiCuMoNb 5	(1.6368) KTA 3211.1, app. A1	q + t, 870–910°C/water +640–660°C	580–620
17 MnMoV 6 4	VdTÜV-Werkstoffblatt 376 (1.8817)	q + t, 880–940°C/air +610–680°C	530–590
X5 CrNi 18 10	DIN 17440 (1.4301) ~AISI 304	1000–1080°C/water, air	—
X6 CrNiTi 18 10	DIN 17440 (1.4541) ~AISI 321	1020–1100°C/water, air	—
X6 CrNiMoTi 17 12 2	DIN 17440 (1.4571) ~AISI 316 Ti	1020–1100°C/water, air	—
X6 CrNiNb 18 10	DIN 17440 (1.4550) ~AISI 347	1020–1100°C/water, air	—
X6 CrNiMoNb 17 12 2	DIN 17440 (1.4580) ~AISI 316 Cb	1020–1100°C/water, air	—

q + t, quenched and tempered.

generally differ for ferritic and austenitic stainless steels.

Ferritic Steels

The as-fabricated condition of the steels are examined with respect to the following.

Microstructural Properties. Specified values of microstructural properties, for example the size and number density of nonmetallic inclusions, segregations, grain size, and hardness similar to the pressure vessel steels described in Sec. 14.2.2.2 of this Chapter, are checked.

Grain Structure. Depending on the heat treatment condition of the steel, for example as listed in Table 7, the microstructure is either quenched and tempered or normalized (Fig. 19).

Due to its influence on strength, ductility, and weldability, the ferritic grain size is restricted [37] to low values (e.g., finer than No. 6). In certain cases of quenched and tempered or HAZ microstructure, the ferritic grains appear as fine bainitic lamellas, and are not easily detectable according to the standardized procedures for grain size measurements. Instead, the former austenite grains are often considered (if clearly apparent) or the bainitic structure is quantified with the line intercept procedures given in the standard procedures for grain size evaluations [38, 39].

Heat-Affected Zone. For the influence of welding on the microstructure and the targets of evaluations, refer to Sec. 14.2.2.2 of this Chapter. In addition to the microstructure, the macroscopic geometry of weld connections of piping is important, in order to avoid local increases in stress intensity due to misfits, local reductions of the wall thickness, or the geometry of the root layer.

Table 8. Applications, service conditions, problem areas, and targets of characterizations

	Ferritic steels	Austenitic steels
Reactor type	PWR, PHWR, BWR, RPV internals, structural parts of fuel elements	
Main use	Main steam line, feed water line, lines for auxiliary systems with large diameters (>DN 100) in secondary systems (e.g., 15 MnNi 6 3, 15 NiCuMoNb 5, 15 Mo 3) auxiliary and monitoring lines with small diameters (e.g., St 35.8, 15 Mo 3, 15 NiCuMoNb 5)	Nuclear auxiliary systems with diameters from DN25 up to DN400, US designs preferentially use unstabilized stainless steels (SSs) European designs use stabilized SSs
Material properties	High strength/ductility at ambient and elevated temperatures Excellent weldability, uniform properties in the base metal, HAZ, and weld metal Excellent failure detection with nondestructive test (NDT) methods (ultrasonic, X-ray, magnetic particle test)	High ductility even at low temperatures, no ductile/brittle fracture transition Excellent weldability, no preheat and post-weld heat treatment required Good failure detection with NDT methods (X-ray, ultrasonic, penetrant test)
Operational conditions		
Operational temperatures (K)	$\leqslant 623$	$\leqslant 623$
Coolant	Feed water PWR, PHWR: low oxygen due to hydrazine or phosphate chemistry BWR: $O_2 < 200$ ppb, depending on the water chemistry	PWR, PHWR and BWR primary coolant as given in Table 3, and coolant conditions as applied for ferritic steels
Coolant pressure (bar)	refer to Table 3	refer to Table 3
Maximum neutron fluence	Not relevant	$\leqslant \sim 10^{22}\,cm^{-2}$

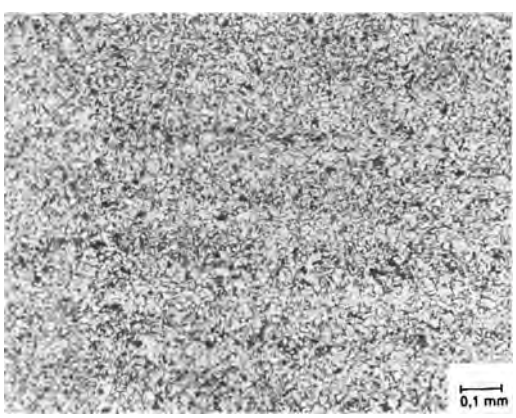


Figure 19. Microstructure of the steel 15 Mo 3 in the normalized condition.

Austenitic Steels

Microstructural Properties. Specified values of microstructural properties, for example the size and number density of nonmetallic inclusions, segregations, precipitates, grain boundaries, and hardness are checked.

The austenitic steels contain sulfidic inclusions and exogenous and endogenous oxidic inclusions. In titanium-stabilized steels, the amount of endogenous oxides is markedly higher due to the affinity of titanium for oxygen, nitrogen and other impurity elements. All stabilized steels may in addition contain coarse carbides or

carbonitrides. The content of sulfidic inclusions is of importance with respect to the corrosion resistance of the material and the weldability. The amount of oxides is of minor influence on the mechanical properties and the weldability of very ductile material, but affects the surface quality and is the cause of frequently detected positive penetrant test indications in machined surfaces or weld edges. Deviations in the hot forming routine (forging with high deformation rates at low temperatures) may cause the formation of numerous small cracks along the stringers of inclusions (Fig. 20).

Because of their excellent ductility, the austenitic stainless steels are susceptible to strain hardening caused by cold deformation after the final solution heat treatment by, for example, machining. As a small amount of cold deformation may accelerate the precipitation of carbides during service at elevated temperatures (>200°C), the strain hardening of weld edges, and also of base material in contact with a corrosive agent, may be restricted to a certain maximum hardness (e.g., 300–350 HV).

A higher degree of cold deformation may cause the formation of martensitic structures with poor corrosion and stress corrosion cracking resistance.

Grain Structure. Except from the δ–γ transition at high temperatures, as will be described below, the austenitic steels are not affected by phase transformations during cooling to ambient temperatures (the formation of martensitic structures in the considered range of chemical composition may occur to a minor extent only at very low temperatures, for instance by cooling with liquid nitrogen). The grain structure results entirely from the hot and/or cold deformation sequence and the recrystallization during the intermediate and final (solution) heat treatments. If the forging results in locally low deformations or the heat treatment sequences (especially the final solution heat treatment) do not allow for complete recrystallization, a mixed grain structure with alternating zones with recrystallized and non-recrystallized zones may result (Fig. 21). As in the non-recrystallized zone, the

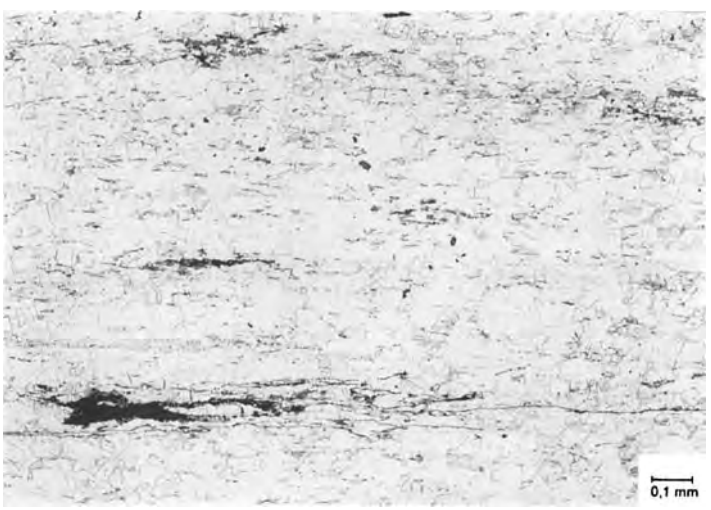


Figure 20. Bands of nonmetallic inclusions (titanium rich oxides and carbonitrides) in the stabilized austenitic stainless steel X6 CrNiTi 18 10. Cracks formed during forging can be seen.

Figure 21. Partially recrystallized microstructure in a forging of stabilized austenitic stainless steel 1.4550.

grain structure is difficult to develop by metallographic etching, and may be considered as having coarse-grained zones, which may restrict the weldability and the resolution of ultrasonic testing. For further investigations, either TEM of thin foil samples or reweld testing as described below may be performed. The reweld test shows in addition to the evaluation of the δ ferrite content the behavior of the non-recrystallized zones during welding, because of the risk of the formation of a coarse-grained HAZs due to critical deformation rates.

Welding and Heat-Affected Zones. During the solidification of molten base or weld metal, the risk of the formation of hot cracks due to the segregation of impurities at the dendritic interface is much higher if the solidification takes place in the f.c.c. γ phase (austenite). Therefore, the chemical composition of the materials is specified within a range where the solidification will occur in the body-centered cubic (b.c.c.) δ ferrite phase followed by a solid–solid transformation to the γ phase. Depending on the cooling rate during welding or the thermomechanical treatment during the fabrication of the semi-finished product, a certain amount of δ ferrite phase may remain in the microstructure. The total amount of possible δ ferrite content is specified to range from 2 vol.% up to about 12 vol.%; at higher concentrations, care must be taken to prevent large networks of δ ferrite in welds because of the risk of low-temperature brittle behavior, corrosion attack of the ferritic phase or the ferrite/austenite interphase, and long-term embrittlement during service at elevated temperatures caused by the formation of the σ phase [40].

The amount of δ ferrite may be evaluated by several methods:

- calculation by means of the chemical composition and the formulas and diagrams given by Schaeffler [41] or DeLong [42];
- electromagnetic measurement of the ferromagnetic δ phase by the Förster— or Fischer—probe; and
- metallographic evaluation (Murakami etchant, Table 9) of the amount of δ

Table 9. Survey of procedures for the microscopic examination and sample preparation of ferritic and austenitic stainless steels

Target or characterization	Light microscopy	Scanning electron microscopy	Transmission electron microscopy
As-fabricated condition	*All materials* Impurities Cracks/flaws formed during fabrication Grain size Grain structure, recrystallization HAZ characterization Weld geometry, hardness *Austenitic steels* Grain size, deformation structures Hardness of machined surfaces δ ferrite content in base metal and HAZ Carbide precipitation in HAZ Hot cracks, by liquations in HAZ	*All materials* Irregularities (e.g., flaws, or inclusions of impurities) in fracture of samples for mechanical testing Structure and composition of contaminants on the surface	*Ferritic steels* Crystal structure and chemical composition of the matrix (microsegregations, concentration profiles) Morphology and chemical composition of carbides (precipitates) *Austenitic steels* Dislocation structures (incomplete recrystallization during hot forming) Chromium-rich carbides on grain boundaries and formation of chromium-depleted zones (sensitization against IGA, IGSCC)
During/after operation	*All materials* Verification of NDT indications Characterization of corrosive attack or failures caused by corrosion *Ferritic steels* Attack by boric acid Environmental assisted cracking *Austenitic steels* Stress corrosion cracking, TGSCC, IGSCC	*All materials* Fractographic characterization of surface attack or cracks caused by the operation Microanalytical identification of characteristic contaminants promoting specific corrosive attack *Austenitic steels* Pitting attack or TGSCC caused by chlorides	*All materials* Service induced changes in the microstructure caused by: – Temperature with respect to mechanical properties (fracture toughness) corrosion resistance (sensitization) – Fatigue strain (dislocation networks)

Sample preparation	*Ferritic steels* Refer to Table 5 *Austenitic steels* Standard metallographic grinding and polishing procedure *Etchant for grain structure* E.g. V_2A etchant: 100 vol.% HCl, $T = \sim 50°C$ 100 vol.% H_2O, $t = 10–30$ s 10 vol.% HNO_3 *δ-ferrite* E.g. Murakami etchant, 90 wt.% H_2O, 10 wt.% NaOH, 10 wt.% $K_3Fe(CN)_6$ *Precipitates on grain boundaries* E.g. electrolytic etching with 90 wt.% H_2O, 10 wt.% oxalic acid, DC voltage: ~40–50 V	*Ferritic steels* Refer to Table 5 *Austenitic steels* SEM: no special procedures needed. On fracture surfaces or surfaces covered with chromium-rich oxides, the scales may be removed without attack of the base metal by a two step cleaning procedure: I: Oxidizing treatment with a mixture of 100 g NaOH and 30 g $KMnO_4$ in 1 l H_2O II: Solution of the oxide by the endox 214 method, refer to Table 3 Repeat steps until desired cleaning effect is obtained	*Ferritic steels* Refer to Table 5 *Austenitic steels: matrix* Standard preparation of TEM samples as given in Table 5 *Precipitates* Standard metallographic preparation Deep etching with V_2A etchant Extraction of the precipitates with a standard replica technique
Optical instruments, magnification range	Refer to Table 5	Refer to Table 5	Refer to Table 5

IGA, intergranular attack; IGSCC, intragranular stress corrosion cracking; TRSCC, transgranular, transgranular stress corrosion.

ferrite phase by comparison with standard series or image analysis.

An experimental estimation of the potential of a material to form δ ferrite may be performed by a local welding of the material under standardized conditions [20] (e.g., remelting of the zone of interest with tungsten inert gas (TIG) welding, ~20 V, ~160 A, ~200 mm min^{-1}) followed by a metallographic or field-metallographic examination or by magnetic measurements.

The remelt test is also used to check the susceptibility of a material to form hot cracks ('liquations') in the HAZs during welding.

This mode of failure during fabrication occurs most frequently in niobium-stabilized stainless steels because the excess niobium, which is not ligated to carbon, is able to form eutectic intermetallics with the matrix [43] (Fig. 22). The risk for the formation of cracks is higher in coarse-grained base materials or HAZ structures (influence of welding), and may be minimized by restrictions on the grain size of the base material (<No. 4), the chemical composition, and the heat input during welding.

Figure 22. Reheat cracks in the HAZ of a weld in a coarse-grained austenitic stainless steel 1.4550 (field metallographic replica technique).

14.2.3.3 Characterization and Evaluation During and After Service

Ferritic Steels

Corrosion. During operation at elevated service temperatures in water with a low oxygen content, the ferritic materials will form a dense layer of oxides (magnetite), which protects the surface from further corrosive attack. At lower temperatures (e.g., during outages) water with a high oxygen content is able to penetrate the shielding oxide layer, resulting in a mould corrosion attack with a low penetration depth and a porous, pustular oxide. During service, a repassivation of the moulded zones will take place, if the oxygen level in the water decreases again to a low level.

In the case of components that operate in contact with the coolant water, the hardness of the materials, especially of the HAZs, should not exceed the limit value of about 350 HV10 [44] in order to avoid failure by hydrogen-induced stress corrosion cracking, which often appears as intergranular cracking of the austenite grain boundaries in the coarse-grained zone of an HAZ (Fig. 23). In direct contact with water containing high oxygen levels (e.g., during commissioning or after outage), the steels, as most of the low-alloyed ferritic steels, may be attacked by environmental assisted cracking (EAC; here strain-induced cracking). The cracks are formed due to the concentration of strain at zones with higher stress intensity (e.g., notches, geometric mismatches) or locally lower yield strength (Fig. 24) [45–47] in the weld metal or the HAZ and in very minor amount in the base materials.

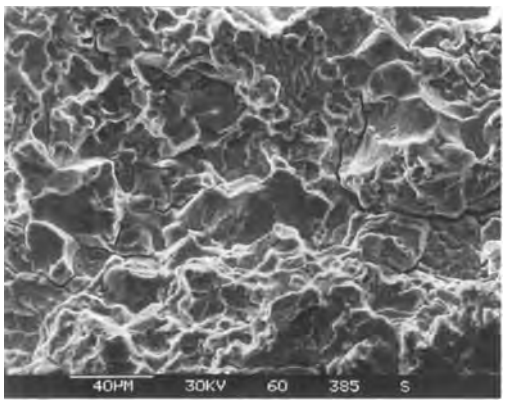


Figure 23. Intergranular and transgranular brittle crack facets due to hydrogen induced cracking in an HAZ of a ferritic steel (15 Mo 3). Crack surface examined by SEM.

Low Cycle Fatigue. Cracks formed due to low or high cycle fatigue mostly occur in combination with EAC, and the mechanical effects are barely detected by fractograph due to secondary corrosive attack. The examination of changes in the microstructure may reveal more information (see Sec. 14.2.2).

Irradiation- and Temperature-Induced Changes in Properties. In general these are not relevant for LWR piping (see Sec. 14.2.2).

Figure 24. Environmentally assisted (strain-induced) cracking in the repair weld and HAZ of a quenched and tempered ferritic tube weld (material 15 NiCuMoNb 5).

Austenitic Steels

The austenitic steels as described in Table 6 are sensitive to contaminants containing high concentrations of chlorides in aqueous solution in the temperature range between about 50 and 150°C, which will cause pitting corrosion or, under the influence of operational or residual stresses, transgranular stress corrosion cracking (TGSCC (Fig. 25) [48]). To avoid failures, strict restrictions on the use of chlorine-containing lubricants and cleaning agents, or other chlorine sources (e.g., sealings containing asbestos) are required. The higher level of oxygen (about 150–200 ppb) and free radicals (e.g., from H_2O_2) in the coolant of boiling water

Figure 25. TRSCC caused by local contamination with chloride-containing impurities in a tube of the austenitic stainless steel 1.4550.

reactors may cause intergranular stress corrosion cracking (IGSCC) in pipe systems which draw high-temperature water from the reactor. This mode of failure is well known to occur in unstabilized stainless steels [49, 50]. Since 1993, some cases have become apparent in titanium-stabilized austenitic steels (1.4541) as well [51, 52].

The cracking occurs mostly in the coarse-grained HAZ of welds and in the weld metal (Fig. 26). TEM examinations of extraction replica show that the cracks mostly follow a path close to the fusion line, where the grain boundaries are decorated with chromium-rich carbides in addition to the carbides resulting from the stabilization (TiC) (Fig. 27). In order to avoid the risk of cracking, restrictions on chemical composition ($C \leqslant 0.04$ wt.%, $Ti/C \geqslant 8$) are required, and care should be taken to machining and welding processes during fabrication (cold deformation of weld edges, heat input during welding). As the austenitic steels have a higher coefficient of thermal expansion, the risk of the introduction of high residual tensile stresses due to welding may be reduced by using special welding techniques (e.g., reduction of welding volume, narrow gap welding) or post-weld thermal or mechanical treatments.

Figure 26. IGSCC in BWR high-temperature reactor water in the HAZ of a tube of the austenitic stainless steel 1.4541.

Figure 27. Chromium-rich carbides along the grain boundaries in addition to Ti-rich stabilizing carbides in austenitic stainless steel.

14.2.4 Materials for Steam Generator Tubing and High Strength Components

14.2.4.1 Survey of Alloys and Data

The nickel-base alloys Inconel 600 and 690 and the high nickel and chromium alloyed

stainless steel Incoloy 800 exhibit excellent uniform and localized corrosion resistance in various environments, and are therefore used for heat exchanger tubes in PWR and PHWR steam generators [53, 54].

The nickel base alloys Inconel X-750 and 718 are used in precipitation hardened condition for components that require high yield strength and high resistance against thermally or irradiation-induced relaxation [55, 56].

The chemical composition ranges of the alloys typical for nuclear application are listed in Table 10 and a brief description of material specifications and typical heat treatment sequences is given in Table 11.

The fields of applications, and service conditions are given in Tables 12 and 13. The procedures for examination are closely related to those of austenitic stainless steels (see Sec. 14.2.3). Specific targets for characterization and procedures for examination are given in Table 13.

14.2.4.2 Characterization and Evaluation in the As-Fabricated State

The microstructure of the materials corresponds very closely to that of the austenitic stainless steels described in Sec. 14.2.3 of this Chapter, and many of the targets for examination are very similar.

Microstructural Properties

The specified values for the microstructural properties, such as the size and number density of inclusions, segregations, grain size, and hardness, is checked.

Most of the targets correspond to those of the austenitic steels. Depending on the carbon content, the materials may contain various large inclusions of carbonitrides (Fig. 28) affecting the surface quality.

Hardness testing is of use to check for the effectiveness of the heat treatment sequence for age hardening. The hardness may reach values up to 500 HV. The resistance of the materials to IGSCC is very sensitive to the heat treatment procedure.

As comparable hardness levels may be achieved by differing heat treatment sequences, care must be taken not to use the hardness as the single criterium to qualify the final product.

Grain Structure and Precipitates

The grain structure of the materials in general corresponds to the recrystallized condition. The ranges for the grain sizes are specified with respect to the intended application. In nuclear applications, for the sake of resistance against IGSCC and low-temperature strength, the grain size is specified to fine values (<No. 4 according to ASTM E 112), whereas high-temperature applications require a coarser grain structure (>No. 3 acc. to ASTM E 112) with respect to increased creep strength.

Depending on the amount of the alloying elements titanium, niobium, and aluminum, all alloys, especially the age hardenable ones, are able to precipitate intermetallic phases in blocky or needle-like incoherent morphology ($Ni_3(Al,Ti)$, or $Ni_3(Al,Nb)$), during heat treatments in the range of 600–800°C, often forming a dense decoration along the grain boundaries (Fig. 29). Titanium or niobium carbides, as well as chromium rich carbides, may also occur in a comparable morphology.

The precipitation of chromium-rich carbides may be accompanied by the formation of chromium-depleted zones along the

Table 10. Chemical composition of the alloys

Chemical composition (wt.%)	C	Si	Mn	P	S	Al	Co[a]	Cr	Cu	Ni	Ti	Fe	Others
Inconel 600	≤0.015 (0.010–0.050)	≤0.50	≤1.0	≤0.025	≤0.015	≤0.50	≤1.0	14–17	≤0.50	Balance	≤0.56	6–10	—
Inconel 690	≤0.015 (0.010–0.040)	≤0.50	≤1.0	≤0.025	≤0.015	≤5.0	≤1.0	27–31	≤0.50	Balance	≤0.56	7–11	—
Inconel 800	≤0.03	0.3–0.7	0.4–1.0	≤0.020	≤0.015	0.15–0.45	≤1.0	20.0–23.0	≤0.75	32.0–35.0	≤0.60	Balance	—
Inconel X-750	≤0.08	≤0.35	≤0.35		≤0.1	0.40–1.0	≤1.0	14.0–17.0	≤0.50	≤70	2.25–2.75	Balance	Nb: 0.70–1.20
Inconel 718	≤0.08	≤0.35	≤0.35	≤0.015	≤0.015	0.20–0.80	≤1.0	17.0–21.0	≤0.30	50.0–55.0	0.65–1.15	Balance	Nb: 4.75–5.50 Mo: 2.80–3.30

[a] For applications in primary systems the cobalt content is restricted (e.g., ≤0.1%).

Table 11. Material specifications and heat treatment sequences

Material	Specifications	Mill annealing	Age hardening
Inconel 600	KTA 3201.1, A2, ASTM B 163-93	980–1030°C/inert air, water	[a]
Inconel 690	ASTM B 163-93	980–1030°C/inert air, water	[a]
Inconel 800	KTA 3201.1, A2, ASTM B 163-93	980–1050°C/inert air, water	—
Inconel X-750		980–1050°C/air[b]	730°C/8 h+ 620°C/8 h[b]
Iconel 718		980–1050°C/air[b]	730°C/8 h+ 620°C/8 h[b]

[a] For improved resistance to sensitization against IGSCC, an additional thermal treatment may be applied, for example ~700°C/several hours.
[b] Typical heat treatment sequences. For individual applications the parameters may be varied in a wide range.

grain boundaries, affecting the susceptibility to IGSCC [57–59] (Fig. 30).

The age-hardening effect is caused by the formation of semicoherent platelets of γ' ($Ni_3(Ti,Al)$) or γ'' ($Ni_3(Nb,Al)$) in the grains. Depending on the heat treatment sequence, the platelets occur in monomodal or heteromodal size distributions (Fig. 31), and a γ'/γ''-depleted zone may be apparent parallel to the decorated grain

Table 12. Applications and service conditions

Application	Alloy	Comments
Reactors	Inconel 600, 690, 800 Inconel X-750, 718	PWR, PHWR, BWR
Components	Inconel 600, 690, 800	Heat exchanger tubes for nuclear steam generators
	Inconel X-750, 718	Bolts, screws, springs in valves, core structures, and fuel elements
Material properties	Inconel X-750, 718	High strength and high resistance against thermally and irradiation-induced relaxation (age hardened)
	All alloys	High corrosion resistance
Operational conditions	See Table 3	
Operational temperatures		
Coolant		
Primary side		See Table 3
Secondary side		Feedwater, ≤ 0.2 ppm O_2, additions of hydrazine or phosphates, enrichment of impurities in the lower zone of a steam generator (phosphate wastage zone)
Coolant pressure		
Primary side		See Table 3
Secondary side (steam)		~65 bar
Maximum neutron fluence ($E > 1$ MeV)	Inconel 600, 690, 800 Inconel X-750, 718	— $\leq \approx 1 \times 10^{22}\,cm^{-2}$

Table 13. Survey of procedures for the examination and sample preparation of Inconel 600, 690, 800, X-750, and X-718

Target	Light microscopy	Scanning electron microscopy	Transmission electron microscopy
As-fabricated condition	Impurities Cracks/flaws formed during fabrication Grain size Grain structure Precipitates and segregations on the grain boundaries Integrity of welds (hot cracks)	Irregularities (e.g., flaws) or inclusions of impurities in the ductile fracture Structure and composition of contaminants on the surface (S!) Impurities causing hot cracks	Details of the microstructure with respect to: – Dislocation arrangement and density – Morphology and chemical composition of precipitates (chromium-rich carbides, intermetallics γ', γ'' phases)
During/after operation	Verification of NDT-indications Corrosive attack (TGSCC, IGSCC, wastage corrosion)	Characterization of cracks caused by the operation (inter- or transgranular)	Temperature or irradiation-induced changes in the microstructure with respect to the formation of precipitates/segregates, carbides, or dislocation arrangements
Sample preparation	Standard metallographic grinding and etching procedure Etchant to reveal: – grain structure: electrolytical etching with 10 wt.% oxalic acid in H_2O, $U \approx 3$ V, $t \approx 15$ s – precipitates [65]: swab etching in HCl, $t \approx 30$ s, cleaning in CH_3OH, etching in a solution of 2 wt.% Br_2 in CH_3OH, $t \approx 10$ s	SEM: no special procedures needed, for removal of oxide layers, see Table 9 for austenitic steels. Deep etching enables examination of coarse intermetallic or γ', γ'' phases, chromium carbides are lost (formation of etch pits)	Matrix: see Table 9, austenitic steels Precipitates: extraction replica technique after deep etching is expected to be applicable (see Table 9 for austenitic steels), but is not applied in general
Optical instruments used and magnification range	See Table 5	See Table 5	See Table 5

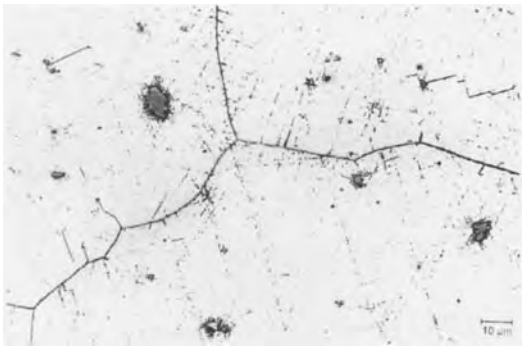


Figure 28. Coarse primary carbonitrides and γ'-decorations of slip bands and dislocations in Inconel X-750 after 1150°C/0.5 h/H_2O + 739°C/16 h/air.

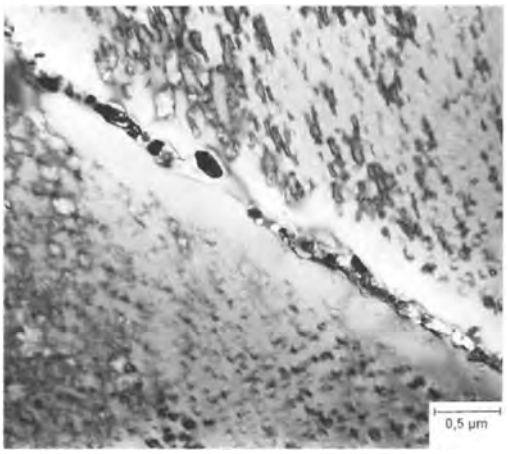


Figure 30. Incoherent precipitates of the phases γ' ($Ni_3(Al,Ti)$), MC (M = Ti, Nb), and $Cr_{23}C_6$, and γ'-denuded zones at a grain boundary of the nickel-base alloy Inconel X-750 in the age-hardened condition (TEM image of a thin foil sample).

boundaries. A cold deformation prior to the age-hardening treatment may result in an inhomogeneous distribution of the precipitates due to the formation of decorations on slip bands and dislocations.

Welding, Rolling, and Annealing

The alloys are more susceptible to the formation of hot cracks during solidification, compared to the austenitic stainless steels. Nevertheless, the weldability is sufficient to perform high-quality seal or repair welds. In age-hardened materials, the microstructure and strength in the weld and HAZ may be changed completely, resulting in poor strength and corrosion resistance in addition to high-tensile residual stresses. Tensile residual stresses

Figure 29. Deep etched structure of the γ' phase at the grain boundary and in the grains of Inconel X-750. Electrolytic etching by orthophosphoric acid (SEM image).

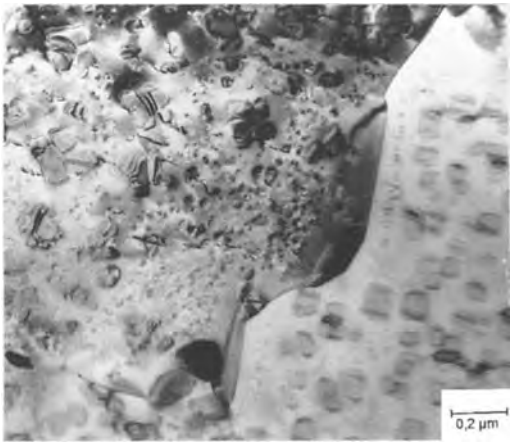


Figure 31. Hardening by semicoherent precipitates in a bimodal size distribution caused by a two-step age-hardening treatment of Inconel X-750 (TEM image of a thin foil sample).

may also promote the formation of IGSCC, and cold deformation processes (e.g., rolling in the manufacture of steam generator tubes from tube sheet) must be applied with qualifications and care.

All materials, especially those with high nickel contents, are sensitive to contaminants containing sulfur (e.g., lubricants) during annealing above temperatures of about 400°C, necessitating a thorough cleaning prior to heat treatments or service.

14.2.4.3 Characterization During and After Service

Corrosion

The alloys Inconel 600 and X-750 are sensitive to primary waterside IGSCC under the influence of tensile stresses (service and residual stresses!), and to the formation of a sensitive microstructure caused by segregation of impurities (e.g., phosphorus and sulfur), and/or the precipitation of chromium-rich carbides (chromium-depleted zones!) at the grain boundaries [53, 59–61] during long-term service (low-temperature sensitization, LTS). Additional thermal treatments may reduce the risk of LTS significantly [62, 63]. The alloys Inconel 800 and 690 under optimal conditions have, from practical experience, proven to be insensitive to LTS [53].

The alloy Inconel 718 is generally used in fuel element structural components where short operational periods reduce the risk of IGSCC. Steam generator tubes may suffer from the so-called phosphate wastage zones caused by the enrichment of degrading slurry in the lower zone of the secondary side of the steam generator [64].

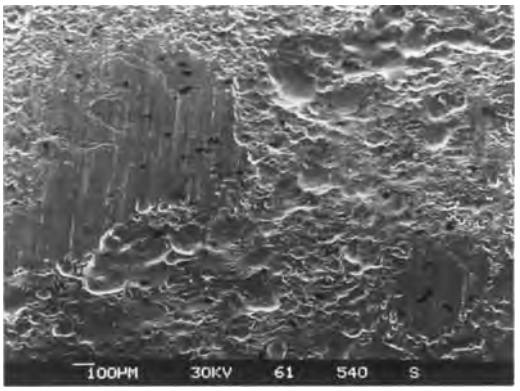


Figure 32. Wastage corrosion attack at the outer surface of a steam generator heat exchanger tube of the austenitic stainless steel Incoloy 800 (SEM image).

The corrosion attack occurs uniformly and may even cause local perforations (Fig. 32).

Fatigue

The age-hardened alloys used for mechanical components (e.g., springs) may fail by high cycle fatigue cracking. The fatigue cracks in coarse-grained materials often exhibit a transgranular fracture with fatigue striations within an intergranular structure (Fig. 33).

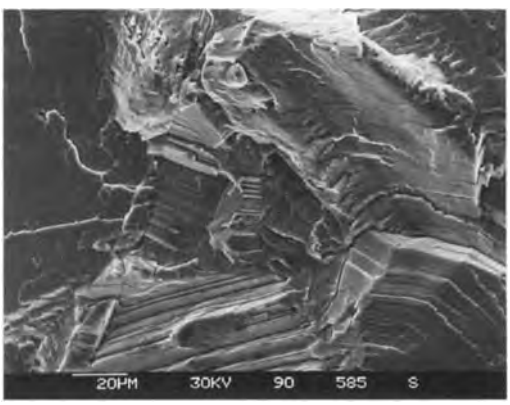


Figure 33. Fatigue crack in a spring of Inconel X-750 (SEM image of the crack surface).

14.2.5 Zirconium-Based Cladding Materials and Hafnium

14.2.5.1 Zirconium-Based Alloys and Hafnium

Zirconium-based alloys as given in Table 14 have been used for fuel claddings and structural parts of fuel elements since the beginnings of nuclear power production, and have proven their efficiency and reliability in many hundreds of thousands of fuel rods. Hafnium has practically the same metallic properties as zirconium but opposite nuclear ones, that is, a high cross-section for absorbing neutrons. Therefore, hafnium is separated from zirconium during the raffination process and used for the fabrication of control rods. The zirconium alloys are usually fabricated from two or three times vacuum remolten ingots of sponge zirconium [66] together with recycled material and the alloying elements tin, iron, chromium, nickel, and niobium. The ingots are hot forged or rolled, extruded, and cold pilgered or cold rolled to their final dimensions in several steps, followed by stress relief or recrystallization treatments (Table 15).

Zirconium (and hafnium) and its alloys are characterized by a phase

Table 14. Chemical composition of hafnium, zirconium, Zircaloy 2 and 4, and Zr-2,5Nb

Chemical composition (wt.%)[a]	Sn	Fe	Cr	Ni	Nb	O	Others
Hf							Hf ⩾ 99.5
Zr							Zr ⩾ 99.85
Zircaloy 2	1.20–1.70	0.07–0.20	0.05–0.15	0.03–0.08		0.09–0.16	Fe + Cr + Ni: 0.18–0.38
Zircaloy 4	1.20–1.70	0.18–0.24	0.07–0.13			0.09–0.16	Fe + Cr: 0.28–0.37
Zr-2,5Nb					2.40–2.80	0.09–0.13	

[a] All alloys may be specified with individual restrictions on the chemical composition (e.g., Sn, Fe, Cr, O, C, Si) with respect to the application.

Table 15. Material specifications and heat treatment range of hafnium, zirconium, Zircaloy 2 and 4, and Zr-2,5Nb

Material	Specifications	Heat treatment range	
		Stress relief	Recrystallization
Hf	ASTM-B 737 ASTM-B 776	~450–650°C/~1–6 h	~650–750°C/~1–6 h
Zr	ASTM-B 352 ASTM-B 353	~450–500°C/1–6 h	~500–650°C/~1–6 h
Zircaloy 2 Zircaloy 4	ASTM-B811	~480–520°C/~1–6 h	~530–61-°C/~1–6 h
Zircaloy 2 Zircaloy 4 Zr-2,5Nb	ASTM-B 351 ASTM-B 353 ASTM-B353	~480–520°C/~1–6 h	~530–610°C/~1–6 h

transformation from the high temperature b.c.c. β structure above about 1000°C, with an $\alpha + \beta$ two-phase interval between ~1000 and ~850°C and the transformation to the hexagonal close-packed (h.c.p.) α-zirconium phase field below this temperature. Because of the anisotropic deformation behavior of α-zirconium, all α-phase materials possess a texture specific for the semifinished products [67–69], and behave anisotropically.

Applications, service conditions, problem areas, and targets of characterization are given in Tables 16 to 18. Because of the general tendency to increase burn-up of the nuclear fuel, more recently the study of the waterside corrosion of Zircaloy claddings has led to the application of SEM and TEM to analyze the oxide scales and the intermetallic precipitates of the cladding material at the end of operation in commercial PWRs, BWRs, and HWRs, see Table 18.

14.2.5.2 Characterization of the As-Fabricated State of Zirconium-Based Alloys

The various alloys in Table 14, in principle, behave similarly, with respect to the metallurgical treatments. Nevertheless, the individual materials possess characteristic differences important for their applications.

Inclusions and Hardness

Depending on the fabrication process and due to the high solubility of nitrogen and oxygen in zirconium, the zirconium alloys are specified to possess very low levels of nonmetallic inclusions. In general these appear as elongated silicon- and zirconium-rich carbides and silicides with lengths in the range 0.5–5 µm [70]. Hardness is affected by grain size and shape, dislocation density, texture, and chemical composition as a consequence of the history of thermal treatment. Thus, with respect to the special problems of interest, hardness measurements have to be performed and analyzed with care, taking account of the type of alloy and its thermal history.

Grain Size and Structure

Due to the high total deformation which has to be applied during fabrication of tubing and sheet, the grain size in the final product is small. In the recrystallized condition, zirconium, zirconium alloys, and hafnium consist of a homogeneous equiaxed grain structure which is easily detectable by optical microscopy with polarized light (Fig. 34). The structure of zirconium–niobium alloys is more complex and may contain, besides the h.c.p. α phase and depending on the thermomechanical history, one or several niobium-containing phases as the β, β', or ω phases [71].

Depending on the required strength, the alloys may be used in the cold-worked and stress-relieved state, or in a partly recrystallized state [72]. According to the deformation and degree of recrystallization applied during the final production steps the grains are more or less elongated in the direction of deformation and may have irregular shapes. Based on the fabrication parameters and the final dimensions the grain size ranges from about Nos. 8–10 (ASTM E-112) for sheet material or tubing with thick walls ($\geq$2 mm) up to Nos. 15–16 for highly cold worked materials. The

Table 16. Applications and service conditions

	Hf	Zr	Zircaloy 2 and 4	Zr-2,5Nb
Reactor type	PHWR	PWR, BWR	PWR (Zircaloy 4), BWR (Zircaloy 2)	PWR, CANDU PHWR, RMBK, VVER
Components	Absorber rods	Linear barrier against failure due to pellet–cladding interaction (PCI)	Fuel cladding tubes, structural elements of the fuel elements, pressure tubes (old CANDU design, Zircaloy 2), coolant channels (Siemens design, Zircaloy 4)	Fuel elements (Zr–1Nb), pressure tubes (RMBK design)
Material properties	High neutron absorption, good corrosion resistance	Low neutron absorption, low sensitivity for PCI cracking	Low neutron absorption, excellent corrosion resistance, high mechanical and creep strength, low sensitivity to hydrogen and delayed hydrogen cracking	Low neutron absorption, excellent mechanical and creep strength, excellent corrosion resistance
Operational conditions				
Operational temperatures (°C)	$\leqslant\sim$350°C	$\leqslant\sim$400°C	360–400°C	$\leqslant\sim$350°C
Coolant	PHWR as given in Table 3	Not in contact with coolant water	PWR, PHWR and BWR as given in Table 3	PWR and PHWR as given in Table 3
Coolant pressure	As given in Table 3	As given in Table 3	As given in Table 3	As given in Table 3
Maximum neutron fluence (cm^{-2}, $E > 1$ MeV)	$\sim 10^{21}$	$\sim 10^{22}$	$\sim 3 \times 10^{22}$–4×10^{22}	$\sim 3 \times 10^{22}$–4×10^{22}

RBMK, uranium-graphite channel type reactor; VVER, water cooled, water moderated energy reactor; CANDU, *Can*adian *D*euterium *U*ranium.

Table 17. Survey of procedures for the microscopic examination and sample preparation of hafnium, zirconium, and zirconium alloys

Target of characterization	Light microscopy	Scanning electron microscopy	Transmission electron microscopy
As-fabricated condition	Impurities (carbides, silicides, etc.) Crack/flaws formed during fabrication Grain size, size distribution Grain structure Hydride orientation (F_N number) Hardness, quality of the weld	Surface quality Surface contamination Hydride concentration (element contrast!) Eutectics on remelt crack surfaces in HAZ/weld	Degree of recrystallization Size characteristics of second-phase particles Chemical composition and structure of second-phase particles or components of the microstructure
During/after operation	Thickness of oxide scale Concentration and orientation of hydrides Cracks formed due to PCI (BWR-specific fission product-induced SCC)	Structure of the oxide scale Irradiation-induced changes in the size and chemical composition of second-phase particles Fractographic characterization of cracks formed during operation (hydride embrittlement, PCI, etc.)	Irradiation-induced changes to the dislocation structure, formation of black dots, c-component dislocations Irradiation-induced changes in the size, structure, and chemical composition of second-phase particles
Sample preparation	Standard metallographic grinding and polishing procedure *Etchants:* *Hydrides:* 50 vol.% lactic acid 50 vol.% HNO_3 0.1 cm^3 HF *Grain area contrast:* swab etching with cotton wool 45 vol.% H_2O 45 vol.% HNO_3 10 vol.% HF Examine at 50:1–500:1 with polarized light	Deep etching to reveal precipitates *Etchant:* 60 vol.% glycerol 30 vol.% Hf 10 vol.% HNO_3	*Matrix:* Standard preparation of TEM thin foil samples as described in Table 5 Method 2 reduces the risk in hydrogen uptake during electrolytical thinning *Precipitates:* Replica technique applied after deep etching as given for SEM Good experience for Zr–Nb alloys

	Grain boundary marking (method does not work on high cold deformed microstructures): Electrolytic polishing: 50 vol.% methanol 43 vol.% ethylene glycol 4 vol.% perchloric acid 3 vol.% H_2O DC voltage: ~25 V Time: 5–10 s Anodic oxidation of grain boundaries: 46 vol.% ethyl alcohol 27 vol.% H_2O 15 vol.% acetic acid 8 vol.% glycerol 4 vol.% phosphoric acid 1.6 wt.% citric acid DC voltage: ~70 V (!) Time: 5–10 min		
Optical instruments	Refer to Table 5	Refer to Table 5	Refer to Table 5
Magnification range	Polarized light: 50:1–500:1 or 10:1–1000:1	20:1–10 000:1	3000:1–100 000:1 Typical range 20 000:1–60 000:1

PCI, pellet–cladding interaction; SCC, stress corrosion cracking.

Table 18. Survey of procedures for the examination and sample preparation of Zircaloy cladding materials after in-pile service

Target	Light microscopy	Scanning electron microscopy	Transmission electron microscopy
As-fabricated condition	Surface condition, grain size	Surface condition and texture Appearance of precipitates at the surface	Details of the microstructure with respect to dislocation arrangement and density Grain size, texture, and condition of grain boundaries Morphology, density, and composition of intermetallic precipitates
After operation in reactor	Thickness, changes in surface condition, state and morphology of oxide scale		
		Oxide morphology, topography of oxide at oxide/metal interface	Oxide microstructure and structure of barrier layer at oxide/metal interface Changes in the microstructure of the cladding alloy following irradiation, e.g., dislocation density, defect clusters, grain size. Morphology, density and composition of intermetallic precipitates after irradiation, particularly composition variations and changes in the precipitate structure
Sample preparation		Cutting of small pieces of cladding at required pin positions in the hot cells, followed by decontamination For analysis of oxide scale dissolution of cladding alloy in solution of 5–10% Br_2 in ethylacetate at 70°C	Microstructure of cladding: Preparation of 3 mm diameter discs about 150 μm thick, dished to give a center thickness of 20–50 μm Electropolishing of the discs in an electrolyte of 15% perchloric acid/85% methanol, temperature in the range −30 to −15°C, polishing voltage: 25 V
Instruments used	Hot cell microscope	Standard SEM with energy-dispersive X-ray analysis Hot cell SEM with energy-dispersive X-ray analysis and a shielded secondary electron detector	Analytical 200 kV TEM instrument attached to a glove box system, and equipped with STEM and SEM modes and energy-dispersive X-ray analysis
Magnification range	10:1–1000:1	20:1–20 000:1	5000:1–250 000:1

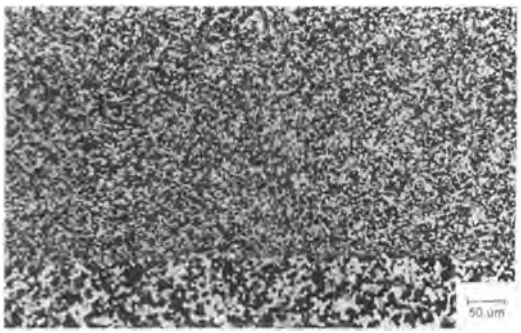


Figure 34. Recrystallized grain structure of a Zircaloy4/zirconium liner fuel cladding tube (polarized light).

method for measuring grain sizes must be carefully chosen because the evaluation by means of standard series given in the standard procedures (e.g., ASTM E-112) may result in considerable deviation when applied with polarized light [73] (Table 17). In stress relieved and partly recrystallized microstructures the evaluation of the grain size requires at least the averaging of three directions (rolling, transverse, and normal directions). For simplicity the grain size is measured frequently following a standardized laboratory heat treatment which results in a recrystallized microstructure. Nevertheless, the deformed microstructure may reveal information on the degree of deformation or the degree of recrystallization if image analysis is applied (Fig. 35).

Hydrides

Hafnium, as well as zirconium and its alloys, exhibit a high affinity for hydrogen. The solubility of hydrogen in the zirconium matrix is a function of temperature, and ranges from very low values at room temperature to about 150 ppm at 350°C [74]. If the hydrogen concentration exceeds the solubility limit, hydrides with compositions between ZrH and ZrH_2 are formed. At lower concentrations at first the f.c.c. δ hydrides ZrH_x ($1.4 \leqslant x \leqslant 1.6$) are formed. With increasing concentration these transform into the tetragonal ϵ-ZrH_{2-x} [75].

With the formation of hydride precipitates the alloy exhibits a ductile to brittle transition in fracture toughness. The transition temperature increases with increasing hydrogen concentration from room temperature to about 200–300°C.

In the absence of stress the hydrides precipitate on lattice planes nearly parallel to the (0001) basal planes [76], and the hydride precipitates reflect the texture. If tensile stresses exist, resulting either from service conditions or from fabrication, the hydrides precipitate perpendicular to the tensile stress axis with a tendency to migrate to locations with stress or temperature gradients (Fig. 36) [77]. The consequence is the formation of high local hydrogen concentrations with brittle behavior. Together with (residual) stresses the locally hydrided zones may form cracks (delayed hydrogen cracking, DHC), for example after cold deformation or welding. The zirconium–niobium alloys with higher strength are more sensitive to DHC compared to the Zircaloys and the unalloyed metal [78]. To minimize the risk of DHC during fabrication the hydrogen concentrations are restricted to low values (zirconium and Zircaloys $\leqslant$ 20 ppm, Zr2.5Nb $\leqslant$ 5 ppm). The influence of the orientation of the platelet-shaped hydride precipitates on the ductility is assessed by determining the orientation distribution of the platelets on metallographic sections by image analysis (F_N number).

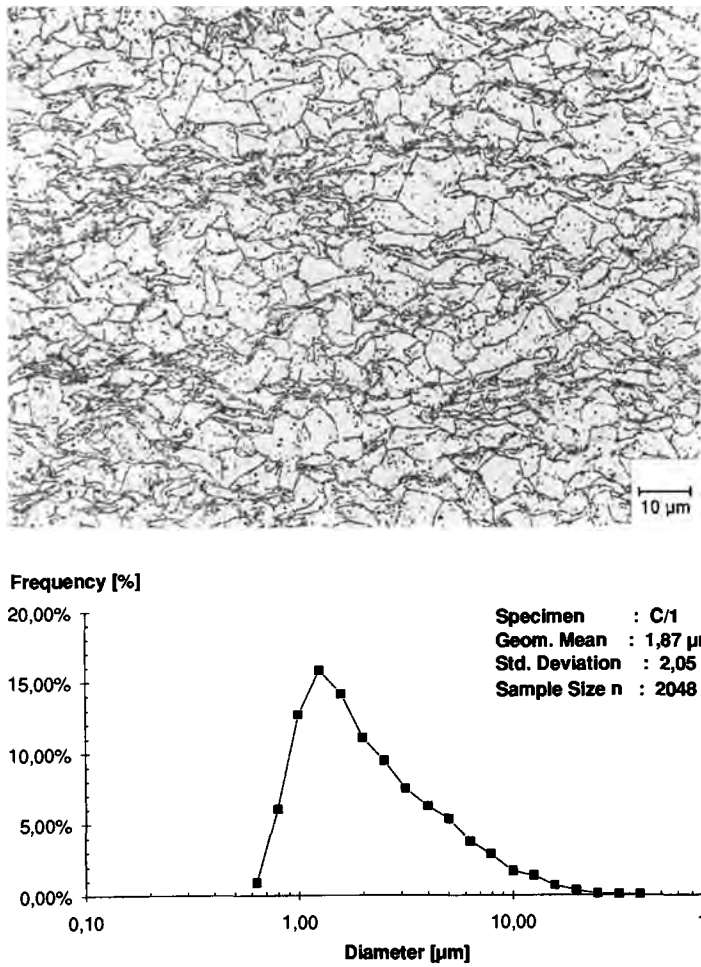


Figure 35. Nonsymmetric grain size frequency plot of a 10% cold-worked Zircaloy 4 sample (optical microscopy of an anodized section).

Recrystallization

Depending on the type of alloy, the materials recrystallize after cold work by annealing for several hours in the temperature range 450–650°C. Deformations of about 10% may produce coarse grains by critical recrystallization. Very high deformations and high annealing temperatures may initiate secondary recrystallization with the formation of very large individual grains [73]. The degree of recrystallization depends sensitively on the final annealing temperature [68] and on texture, and affects the behavior under irradiation in the matrix. Therefore, the ratio between recrystallized grains (low dislocation density) and only-recovered grains (high dislocation density) is determined by TEM (Fig. 37).

Second-Phase Particles (Intermetallics)

The alloying elements iron, chromium, nickel, and niobium are soluble in the high temperature β-zirconium phase. During transformation to the α phase these elements form intermetallic precipitates

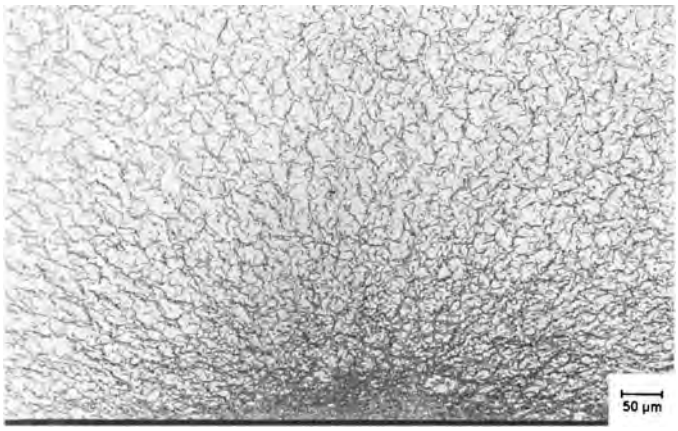


Figure 36. Enrichment of hydride precipitates due to local cold deformation in Zircaloy 4 (optical micrograph).

(Table 17) [79–81]. In Zircaloys the rate of cooling through the $(\beta + \alpha)$ two-phase field determines the density of their nuclei, and further annealing in the high α-phase field their growth [82, 83]. The size distribution of the intermetallic precipitates governs the corrosion behavior during service [84]. The optimized size distribution depends on the water chemistry of the reactor, and must be adjusted individually by modified fabrication routes. Hence the measurement of the size distribution and number density of the precipitates is an essential means for the prediction of the corrosion behavior of fuel rod cladding and fuel element components in service. The sizes of the precipitates occur most frequently in the range 10 nm up to nearly 0.1 µm, with particle densities in the range 1–2 particles per square micrometer of examined area (number density), equivalent to approximately 6–10 particles per cubic micrometer. Individual particles may attain diameters up to 5 µm. The majority of the particles have the structure of the Laves phase $Zr(Fe,Cr)_2$ with ratio $Fe:Cr \approx 2:1$, with a small portion of the precipitates with higher iron content depending on the exact alloy composition. The quality control routines are performed by TEM or SEM [85] (see Sec. 14.2.5.3).

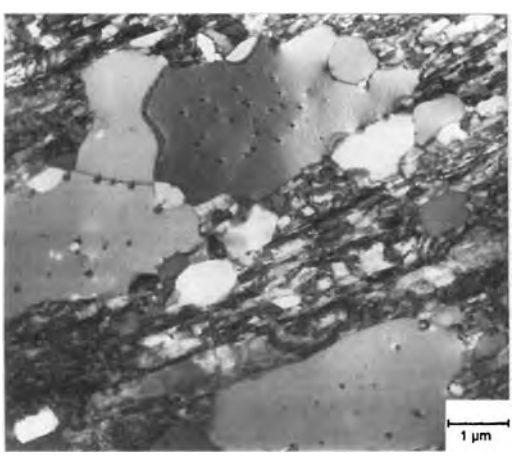


Figure 37. Partially recrystallized microstructure of Zircaloy 4 (TEM image of a thin foil sample).

Welding

Zirconium and its alloys possess an excellent weldability with TIG, laser and electron beam methods (Fig. 38). The affinity of zirconium for oxygen and nitrogen requires very pure protective (argon) atmospheres during welding. Insufficient gas purity results in the formation of tarnish colors by the solvation of oxygen or nitrogen, and may markedly increase

Figure 38. TIG weld on a longitudinal welded tube of Zircaloy 4.

the hardness in the weld metal and the HAZ. This reduces fracture toughness and corrosion resistance, provokes residual stresses, and increases the risk for cracks by DHC after welding.

The melting interval of zirconium alloys is around 1800–1900°C. At these temperatures the elements tin, iron, chromium, and nickel have relatively high vapor pressures, and during welding this is sufficient to reduce their concentrations in the molten alloy. Thus the composition of the welded metal should be compared with that of the base alloy by electron probe microanalysis.

The large range of melting temperatures of the zirconium alloys in combination with surface impurities may cause the formation of hot cracks in the HAZ of the weldings. These appear as intergranular cracks on the β-grain boundaries of Zr2.5Nb. Low-melting phases on the crack surface may be detected by electron probe microanalysis (Fig. 39).

14.2.5.3 Characterization After Service and Irradiation Damage

In addition to the effects of stresses and temperature during reactor operation the fast neutron flux induces changes in the microstructure of the alloy which affect its mechanical properties and corrosion resistance. The most important targets for microscopic examination after service are as follows, Table 18.

Hydrides

During reactor operation the surface of the components of the fuel elements is very slowly oxidized. The oxygen is released by the dissociation of the coolant water. Depending on the coolant chemistry (high or low oxygen concentrations) a certain portion of hydrogen (between 5 and 15%) is picked up by the alloy [86]. During in-pile life the hydrogen concentration in zirconium and Zircaloys may reach local concentrations up to 2000 ppm at special positions due to stress concentration, or temperature gradients.

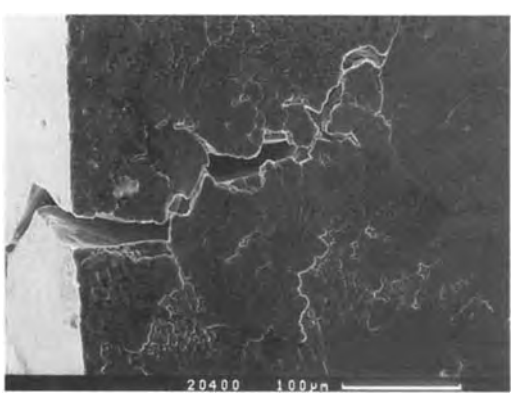


Figure 39. Reheat crack in the HAZ of a Zircaloy 4 weld caused by local impurities (SEM image of the near-weld surface of a section).

Dislocation Structures

After an incubation fluence the vacancy–interstitial pairs (Frenkel pairs) produced by the fast neutron flux at temperatures below 450°C precipitate and form dislocation loops on $\{10\bar{1}0\}$ planes [87–90] (a-type dislocations). The diameter of these loops depends on the irradiation temperature, and they start to become visible by TEM at a fluence $\Phi \sim 10^{19}\,\mathrm{cm}^{-2}$ for $T = 300°\mathrm{C}$ and $E > 1\,\mathrm{MeV}$ as black dots with about 10 nm diameter. Their density increases with the fluence, and seems to saturate at a fast flux of $10^{20}\,\mathrm{cm}^{-2}$. The regular arrangement of the dislocation loops causes a wavy deformation of the thin foil TEM specimens, and appears as a sequence of bright and dark striations ('corduroy' contrast) [91] (Fig. 40).

Depending on texture, the formation of dislocation loops causes irradiation-induced growth of the material. Growth is usually positive along directions normal

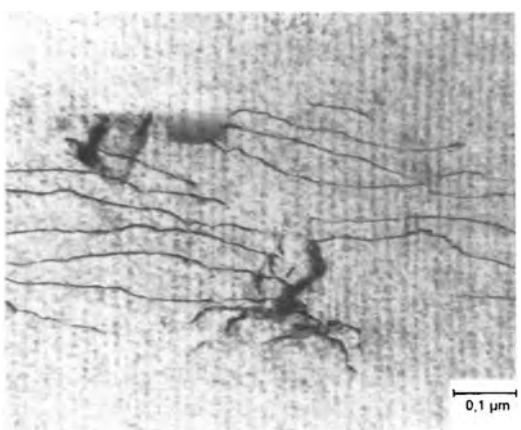


Figure 40. Bands of $\{10\bar{1}0\}$ dislocation loops (black dots, 'corduroy' contrast) in Zircaloy 4 irradiated at 300°C to a fast fluence ($E > 0.8\,\mathrm{MeV}$) of $1.1 \times 10^{22}\,\mathrm{cm}^{-2}$ (TEM image of a thin foil sample).

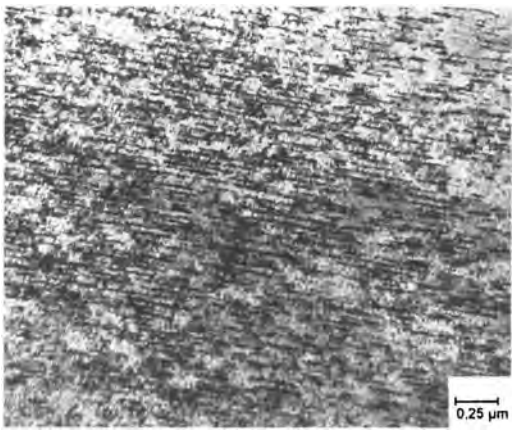


Figure 41. c-component dislocations (black striations) formed by irradiation of Zircaloy 4 at 300°C to a high fluence ($E > 0.8\,\mathrm{MeV}$) of $1.1 \times 10^{22}\,\mathrm{cm}^{-2}$ (TEM image of a thin foil sample).

to $\{11\bar{2}0\}$ and $\{10\bar{1}0\}$ planes and negative normal to (0001) basal planes [92]. This results in a growth in the axial direction of the rod for which the (0001) texture factor (fraction of [0001] poles oriented in axial direction) is very low. At fluences above $3.5 \times 10^{21}\,\mathrm{cm}^{-2}$ ($E > 0.90\,\mathrm{MeV}$) and temperatures ~300°C, new dislocations with Burgers vectors containing a c component are formed in addition to the a-type loops in recrystallized grains [88, 93] (Fig. 41). As a consequence of the formation of c-component dislocations and their increase in density, the rate of irradiation-induced dimensional changes increases after a preceding saturation of the growth [88, 94, 95].

Precipitates (Intermetallics)

Neutron irradiation changes the structure, the chemical composition, the size, and the density of the intermetallic precipitates in Zircaloy [88, 96–98]. Zircaloy 4 contains precipitates of zirconium with iron and of zirconium with iron and chromium. In

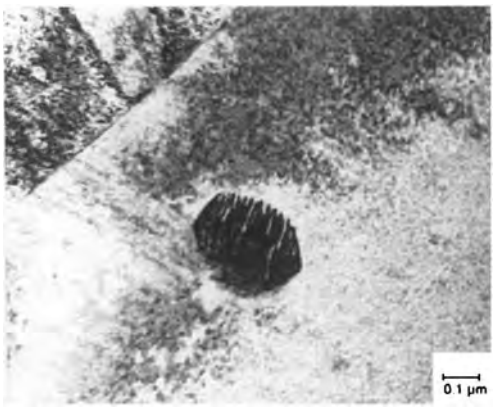


Figure 42. Partial dissolution of a zirconium–iron intermetallic Zr_3Feprecipitate in Zircaloy4 due to irradiation at 300°C to a high fluence ($E > 0.8$ MeV) of 1.7×10^{22} cm^{-2} (TEM image of a thin foil sample).

particular, one observes:

(1) During irradiation in the temperature range 290–310°C the iron goes into solid solution in the matrix, and the precipitates may form lamellas (zirconium–iron type) (Fig. 42), or become amorphous (zirconium–iron–chromium type) and dissolve completely at higher neutron doses.

(2) The density of intermetallic precipitates of the zirconium–iron–chromium type remains approximately constant, but three effects occur [99]:
   (a) The initial composition of the precipitates with ratio Fe:Cr ≈ 7:3 transforms into a spectrum of different compositions ranging from Fe:Cr = 1:0 to Fe:Cr = 0:1, with the majority of intermetallic precipitates in the range 1:9 < Fe:Cr < 4:6.
   (b) Within individual precipitates a radial variation in composition can be found, most frequently with an iron-rich core, as in the example of Fig. 43, with the associated electron dispersive X-ray (EDX) analysis showing the variation of the ratio iron : chromium beween large precipitate, amorphous zone, and small precipitate.
   (c) The original crystalline populations of precipitates transforms to an increasingly amorphous state, an exception to this being the iron-rich precipitates, which remain crystalline. Particles can be found with a crystalline iron-rich core and an amorphous iron-depleted rim. An example of this is shown in Fig. 43 along with the associated diffraction patterns.

At temperatures of 370°C the zirconium–iron–chromium-type intermetallics do not change much. The Fe:Cr ratio remains stable, and some growth may occur. As the precipitates could govern the corrosion behavior of the alloy, depending on the service temperature, a change in the corrosion rate may occur during service up to high fluences, especially under BWR conditions [100].

On cross-sectional specimens of the cladding wall the transition zone between the oxide and the metal matrix can be investigated. Of particular interest is the variation in the type and distribution of the intermetallic precipitates as a function of distance into the metal from the oxide–metal interface.

Electron Microscopy and the Study of Corrosion

Zirconium-based alloys possess relatively good corrosion resistance in water and steam up to 400°C under operating

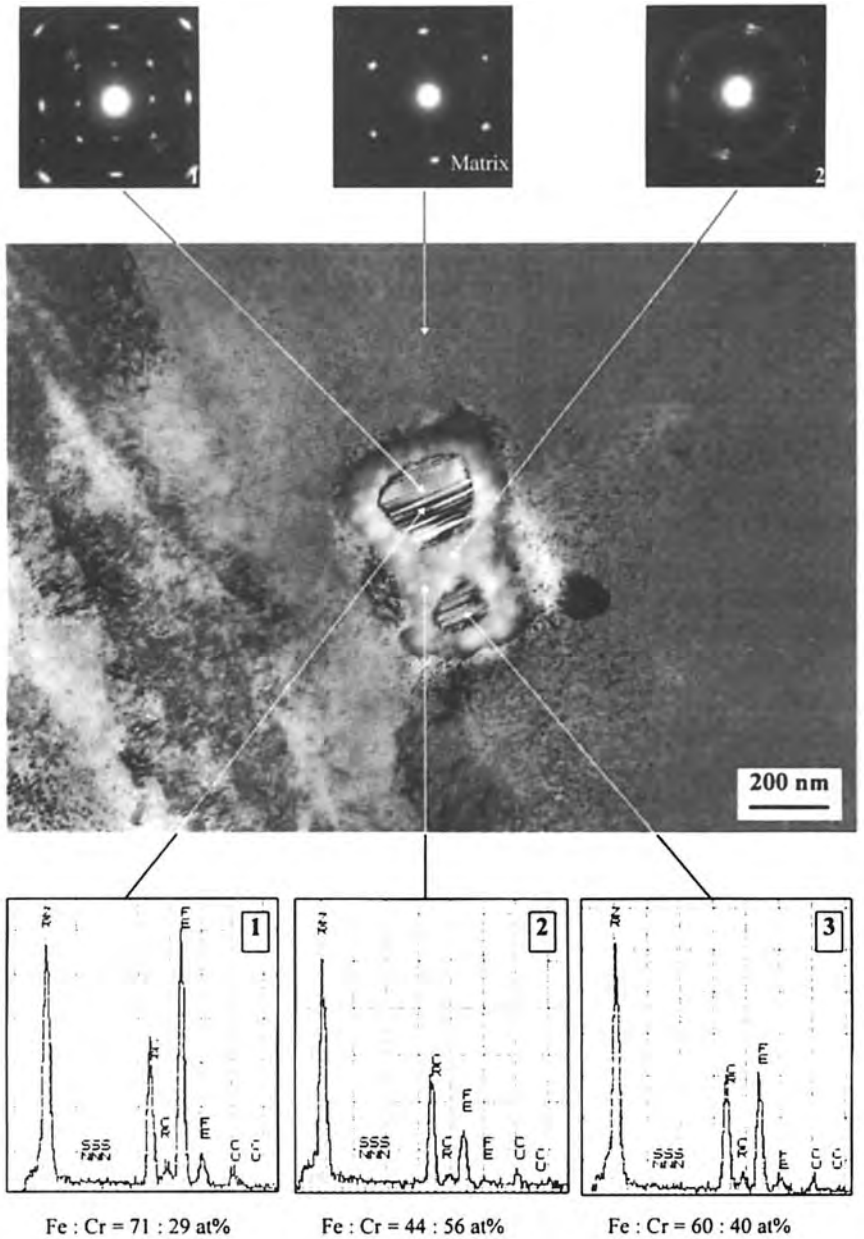


Figure 43. TEM micrograph of two closely spaced $Zr_2(Fe,Cr)$ precipitates in Zircaloy 4 of a cladding section near the cool end of a fuel rod (about 320°C) after four reactor cycles. Note the amorphous zone around the crystalline precipitates as indicated by the selected area diffraction patterns. The lines of Fe and Cr in the local EDX spectra show that Fe is lost from the precipitates and diffuses from the amorphous zone into the α-Zr matrix.

conditions in PWRs, BWRs, and HWRs (see Tables 3 and 16). Nevertheless, with the development of high burn-up fuels and increasing in-pile operating times, corrosion becomes a lifetime-limiting factor for the fuel rods. There is a very abundant literature on the corrosion of zirconium-based alloys (a recent discussion of the state of the art and the related problems with still many open questions in this field has been published by the IAEA [101]). During uniform post-transition corrosion of Zircaloy 4 and Zr2.5Nb the oxide forms a relatively thin (about 2–3 μm) oxide layer, called 'barrier layer', which adheres firmly to the alloy but is periodically replaced and determines the rate of corrosion. Hence, among other fields of parameters, research has concentrated primarily on the analysis of this barrier layer in order to understand the factors which govern the corrosion mechanism. By dissolving the alloy substrate of fuel claddings after irradiation, the morphology and structure of the in-pile grown oxide scale can be analyzed by SEM and TEM [102]. However, it appears that also the detailed structure of the alloy phase close to the oxide–metal interface has to be taken into consideration in order to understand the effects of alloy composition on the post-transition corrosion rate in PWRs [103].

The preparation of TEM specimens as transverse sections across the oxide–metal interface of irradiated cladding material has been performed by grinding and ion etching [104]. Because of the stress gradients existing across the oxide–metal interface in the irradiated bulk material, cracks usually develop in the oxide phase of the TEM specimens parallel to the oxide–metal interface during thinning. This makes successful thinning difficult and impedes the interpretation of the structures.

14.2.6 Structural Materials for Fast Breeder Reactors and Fusion Devices

14.2.6.1 Material Selection and Operating Conditions for Cladding and Wrapper Materials of Fast Breeder Reactor Fuel Elements

The FBRs of today have in common the choice of liquid sodium as the cooling medium with core inlet and outlet temperatures ranging typically from 380 to 560°C. This determines the temperature range of operation for core structural materials such as cladding materials for fuel pins and wrappers of fuel bundles from about 380°C up to 620°C and 580°C, respectively. In exceptional cases hotspot temperatures up to 680°C are to be expected in cladding materials. Hence, corrosion resistance and high-temperature mechanical properties, especially the time-dependent creep rupture strength data and structural stability, are of primary importance for a first material selection.

High-temperature resisting austenitic stainless steels of the type 18CrNi13 (AISI 316) and modifications were the first choice for cladding materials in many FBR projects. Also, high nickel-bearing, precipitation-hardened austenitic alloys of type PE 16 and Inconel 706 and nickel-based alloys such as Inconel 625 have been used. Today, after more than 20 years experience under very high neutron fluences of up to 130 dpa, the

titanium-stabilized austenitic stainless steels of Type 15CrNi15 (DIN 1.4970) in modified version (AIM 1) and the high-nickel austenitic alloy PE 16 are the most promising choices for cladding materials. They should achieve a target burn-up of 15–20% and a displacement dose of 150–200 dpa [105]. Limitations can possibly be expected from irradiation-induced swelling and creep in the austenitic steels at high neutron doses, which lead to unacceptable dimensional changes of fuel elements and to strong ductility losses of the PE 16 alloy.

For wrapper materials with a lower maximum operational temperature, the group of 9–12% CrMoV(Nb) steels has been selected because these alloys possess adequate strength, are sufficiently corrosion-resistant, exhibit low swelling under irradiation, and show no serious embrittlement problems under irradiation [106]. Experience with austenitic stainless steels of types AISI 304 and 316 and other alloys were less encouraging because of strong swelling and coupled irradiation creep.

Table 19 gives typical compositions of cladding and wrapper materials for FBRs, and in Table 20 the optimum thermomechanical treatments for three candidate materials are compiled.

14.2.6.2 Materials for First-Wall and Blanket Structures in Fusion Devices

The first wall of the next fusion device, ITER, and a future DEMO reactor, will be the component with the highest neutron wall loading and—together with the divertor—also have the highest thermal loading. Therefore, the primary selection criteria are favorable thermophysical properties, adequate mechanical and corrosion properties, high radiation resistance, and a low radiation-induced activation. In addition, the tritium pick-up and permeation are of importance. In Table 21, typical operational conditions for the first wall and the divertor of the ITER and a general outlook for the expected loading conditions of a future DEMO reactor are summarized [107]. The first wall is exposed to plasma particles, electromagnetic radiation, and high-energy 14 MeV neutrons. The plasma–wall interaction leads to surface effects such as physical and chemical sputtering and a high thermal wall loading, which is—in combination with the pulsed mode of operation—a major source of thermal fatigue. In addition, plasma disruptions can lead to surface heat loadings which can even exceed 50 MJ m^{-2} in time steps of less than milliseconds. In order to prevent melting and evaporation of the metallic structural materials, protection of the first wall by beryllium, carbon, or tungsten (for the divertor) is foreseen.

The neutron wall loading is characterized by the high energy of the neutrons (14 MeV). Whereas the calculated defect rate of displacements induced by elastic interaction between neutrons and lattice atoms is well in the range of values known from fission reactors, the rate of inelastic reactions increases strongly with the increasing neutron energy. This could have drastic consequences on radiation effects such as helium and hydrogen embrittlement, which are induced by the formation of hydrogen and helium through (n, p) and (n, α) transmutation reactions. Also, the radiation-induced activation is increased by the 14 MeV neutrons.

Table 19. Typical chemical compositions of cladding and wrapper materials for FBRs

Type	Alloy	Fe	Ni	Cr	C	Mo	Mn	Nb	Al	Ti	Si	Other
Austenitic	AISI 316	Balance	13.7	17.3	0.05	2.26	1.64	—	—	—	0.56	
stainless	M316	Balance	13.8	17.3	0.05	2.43	1.78	—	—	—	0.53	
steels	FV548	Balance	11.8	16.5	0.11	1.44	1.14	0.92	—	—	0.35	
	AISI 304	Balance	9.35	18.5	0.07	0.02	1.55	—	—	—	0.48	
	DIN 1.4970	Balance	15.2	14.9	0.10	1.24	1.75	—	—	0.48	0.40	
	DIN 1.4981	Balance	16.0	15.7	0.06	1.78	1.37	0.72	—	—	0.44	
Precipitation-	Nimonic PE 16	Balance	43.4	16.5	0.08	3.15	0.01	—	1.20	1.27	0.01	B = 0.0032
strengthened	Inconel 706	Balance	41.5	16.1	0.03	—	0.10	2.95	0.27	1.66	0.01	B = 0.006 (max.)
alloys	Inconel X-750	7.81	Balance	15.4	0.045	—	0.09	10.6	0.68	2.45	0.07	
	Inconel 625	4.36	Balance	21.9	0.02	9.17	0.09	3.35	0.29	0.28	0.15	
Ferritic–	EM10	Balance	0.18	8.76	0.105	1.05	0.48	—	—	—	0.37	N = 0.024
martensitic	HT9	Balance	0.47	12.0	0.020	1.03	0.50	—	—	—	0.41	W = 0.5, V = 0.32
steels	9Cr–1Mo	Balance	0.09	8.61	0.081	0.89	0.37	0.07	—	—	0.11	V = 0.21
	DIN 1.4914	Balance	0.85	11.0	0.15	0.70	0.70	0.35	—	—	0.40	V = 0.40

Table 20. Recommended typical thermomechanical pretreatment for selected cladding and wrapper materials

Materials group	Type of pretreatment
Austenitic stainless steels Type X 10 CrNiMoTi B 15 15 (DIN 1.4970 mod., AIM 1)	Solution heat treatment (1100–1150°C/10 min) + cold working (20–25%)
Precipitation strengthened alloys Type nimonic PE 16	Solution heat treatment (1075–1100°C/5 min) + ageing (700–720°C/16 h)
Martensitic steels Type X 18 CrMoVNb 12 1 (DIN 1.4914, FV 448)	Austenitization annealing (1050–1075°C/30 min/air) + tempering (740–760°C/1 h/air)

For the ITER project, where only a limited level of radiation damage will be accumulated, the main concerns about material degradations under 14 MeV neutrons are the radiation hardening induced by defect clusters and precipitates at the lower end of the operational temperature range, and the in situ thermal fatigue under irradiation. For the future DEMO reactors, where the operational temperatures and the accumulated neutron damage will be much higher, it is expected that the effects of helium embrittlement, the swelling, and the irradiation creep will play the dominant roles in determining the material behavior [108, 109].

For the ITER project the austenitic stainless steel 316LN has been selected as a candidate structural material, mainly because of the very well established database and the low neutron exposure to be expected. For the long-term aim of a DEMO reactor the European Fusion Technology Programme has included the class of 9–12% CrMoV(Nb) conventional ferritic–martensitic steels, the 8–10% CrWVTa low-activation ferritic–martensitic steels as alternatives, the class of vanadium-based vanadium–chromium–titanium alloys and ceramic composite materials of the type SiC/SiC. Typical compositions of all the alloys are given in Table 22 [110].

Table 21. Typical loading conditions of the first wall and divertor components in the ITER [107]

Parameter	First wall	Divertor
Surface heat flux ($MW\,m^{-2}$)	$\leqslant 0.5$	5–15
Nuclear heating ($MW\,m^{-3}$)	50	
Average neutron wall loading ($MW\,m^{-2}$)	1	0.6
Integral neutron wall loading ($MW\,a\,m^{-2}$)	1–3	0.6–1.8
Defect rate ($dpa\,s^{-1}$)	3×10^{-7}	1.8×10^{-7}
Pulse length (s)	1000	1000
Dwell time (s)	1200	1200
Total number of pulses	4×10^4	4×10^4
Coolant pressure (MPa)	3	3
Coolant temperature (°C)	$\leqslant 50$	$\leqslant 50$

Table 22. Typical chemical composition of materials for first wall and blanket components (wt.% unless indicated otherwise (wt. ppm))

Type	Alloy	Fe	Ni	Cr	C	Mo	Mn	Nb	Al	Ti	Si	Others
Austenitic stainless steel	316 LN	Balance	12.34	17.4	0.02	2.3	1.82	—	—	—	0.46	Cu = 0.2 Co = 0.17 N = 0.06 P = 0.03 S = 0.01 B = 10 ppm
Ferritic–martensitic steels	MANET II	Balance	0.65	10.3	0.11	0.58	0.85	0.14	0.12	—	0.18	P = 0.005 S = 0.004 V = 0.19 B = 70 ppm N = 0.03
	F82H mod.	Balance		7.66	0.09	W 2.0	0.16	Ta 0.02	—	0.01	0.11	V = 0.16 N = 0.005
Vanadium-based alloy	V–4Cr–4Ti	V Balance	—	4	—	—	—	—	—	4	—	N = 0.008 O = 0.036 C = 0.008
Ceramic fiber-reinforced composite	SiC/SiC		—	—	30.1	—	—	—	25 ppm	—	65.7	N = 0.06 O = 4

14.2.6.3 Special Simulation Techniques to Study Radiation Damage

With the development of materials for FBRs and fusion devices, so-called simulation techniques to study different aspects of radiation damage were introduced. Electron irradiation—partially performed in HVEMs—are essentially applied to study the in-situ development of different microstructural components such as dislocations, defect clusters, precipitates, and other properties as indicated in Table 23. Heavy ion irradiation is—for example—extensively used to study high-fluence phenomena such as void formation, precipitate stability, and resolution, as well as ordering/disordering effects. The specific advantage of light ion irradiation is its relatively large penetration depth so that

Table 23. Properties examined by light microscopy, SEM and TEMafter simulating fast neutron damage in FBRs and Fusion Devices

Light microscopy	SEM	Analytical SEM	TEM	Analytical TEM
		EDX/WDX[a]	Bright/dark field techniques and diffraction	EDX/EELS[b]
Unirradiated material				
Grain size, structures and substructures	Grain size and structure	Chemical analysis of matrix, inclusions, and precipitates	Dislocations	Chemical analysis of matrix, inclusions, and precipitates
Inclusions	Inclusions		Density and ***B*** vector determination, Arrangements	
Primary and secondary precipitates	Primary and secondary precipitates	Segregation phenomena	Precipitates: Phase identification, orientation and coherency	Segregation
Fracture analysis and crack formation	Fracture analysis and crack formation		GB structures	
Irradiated material				
Grain size, structure and substructures	As above	As above	Microstructure development under irradiation: Voids, Loops, Helium bubbles, and Precipitates	
Inclusions and precipitates				
Coarsening/dissolution phenomena				
Recrystallization				
Fracture analysis and crack formation				

[a] WDX mainly for irradiated materials.
[b] EELS for light elements.

combined mechanical loadings and irradiation can be applied in in situ tests. This has made the goal of cyclotron irradiation in two-beam constellations very attractive, especially in the field of fusion materials research. All these simulation experiments have the advantage that the control of irradiation conditions such as temperature, flux and fluence, and other parameters is more precise than in very complicated in situ reactor irradiation. Of course, some uncertainties have to be taken into account when the results of such simulation studies have to be transferred to the real situation in a reactor system. To mention only few among many points of concern: the different energy transfer to the lattice during a primary knock-on event, the additional effect of implanted ions on the material, and the performance of experiments in different damage rate regimes. But in general the results of such studies have essentially contributed to the development of materials.

Special problems arise when intensive and quantitative microstructural investigations by TEM have to be performed on samples irradiated with ions. In Fig. 44 the problem is manifested for the case of heavy ion irradiation. Dependent on the energy of the projectile, the radiation damage and implantation profiles are very narrow and at different depths. Dependent on the depth of the damaged zone the preparation of samples for TEM investigations have to be made according to a scheme shown in Fig. 45. For low-energy particles the damaged zone is near the surface, so that an electrolytic back-thinning normally leads to a good sample for TEM. For higher energies when the maximum damage zone is placed far in the bulk of a sample, a selective depth preparation by mechanical, electrolytic, or ion beam thinning techniques has to be adapted, before a back-thinning can again be applied (lower process row in Fig. 45). The problem here is to control the material removal to an accuracy of less than 0.25 μm. For the case of mechanical material take-off by vibratory polishing, this process is controlled by weight-loss measurements and/or by microhardness indentations. Two examples for such investigations in austenitic steels are shown in Figs. 46 and 47, where a direct recrystallization of material under heavy ion irradiation in competition with void formation and sweeping of grain boundaries is being observed by TEM.

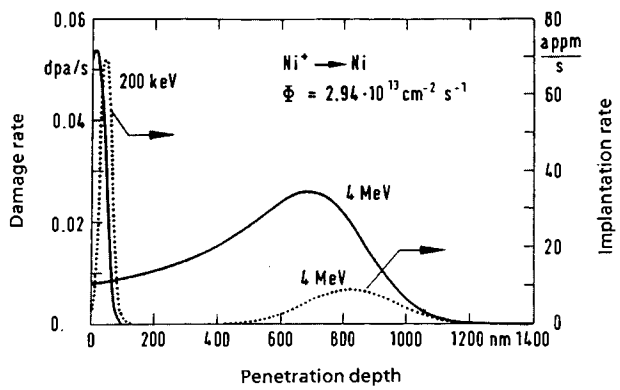


Figure 44. Damage rate and implantation profile in iron under irradiation by nickel ions.

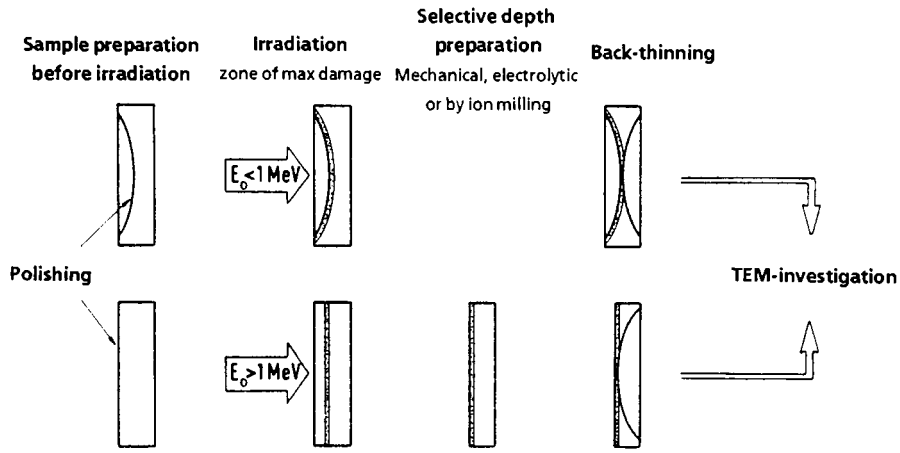


Figure 45. Steps in sample preparation for targets irradiated with low ($E < 1$ MeV) or high energy ions.

A more effective but experimentally very difficult method is to try a cross-sectional preparation of samples so that the whole penetration depth of ion-implanted specimen can be made visible. A method which has recently been developed to investigate the problem of radiation damage and helium bubble formation in ceramic materials for fusion applications is shown in Fig. 48. This method has been used for SiC material irradiated with 200 keV helium ions. After forming a sandwich of several SiC platelets, two of them are implanted with helium, and a transversal cut is polished by inert gas (argon) ions. The thin sections of irradiated samples clearly show in a cross-sectional view the damaged zone with the formation of black dot damage (Fig. 49).

In summary, the simulation irradiation together with a highly developed TEM technique allow the optimum information to be gained about the development of radiation damage that is dependent on dose, temperature, penetration depth, and other irradiation parameters. Very different problems associated with the radiation damage phenomena can be investigated in this way.

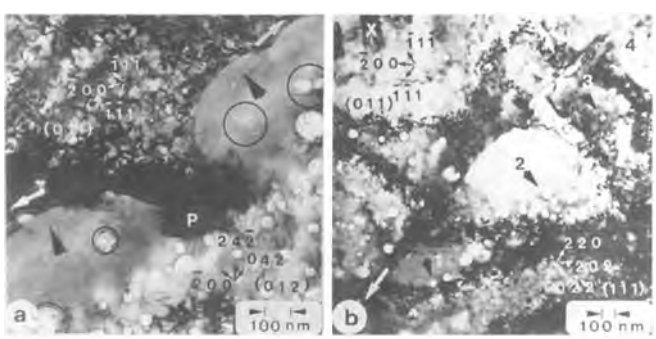


Figure 46. Typical features at the initial grain boundaries in an austenitic stainless steel (DIN 1.4970) during irradiation with 46 MeV nickel ions. Arrows mark the position of the grain boundaries, arrowheads show the direction of migration, and deeply etched voids are encircled.
(a) Migration of the grain boundary pinned by the precipitate P.
(b) Nucleation of grains at the grain boundary.

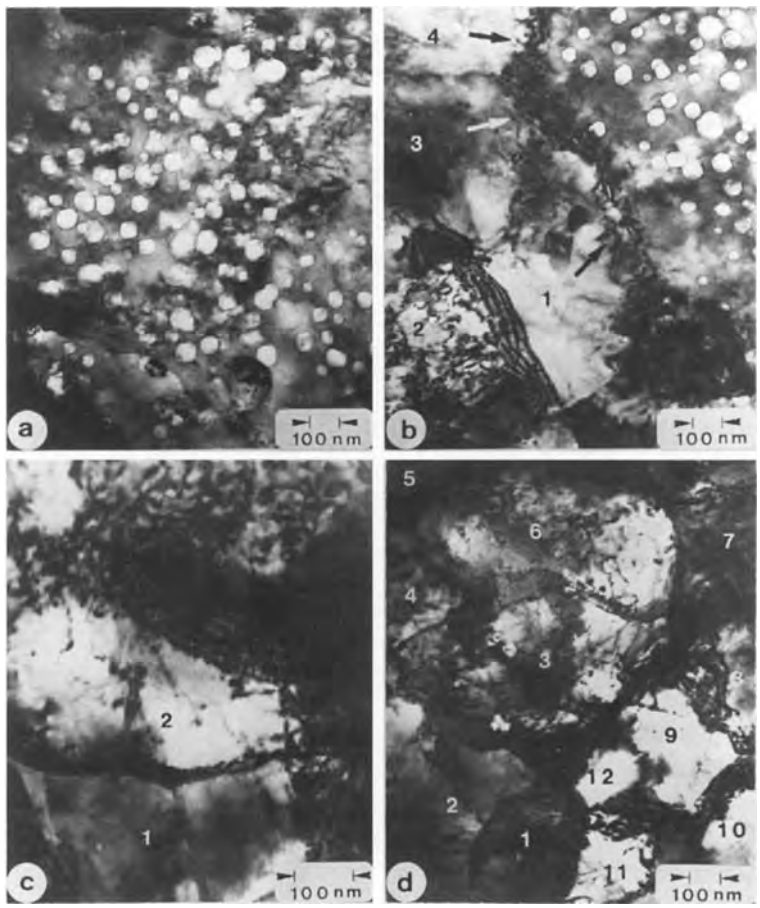


Figure 47. A typical example of interdependency between void formation and recrystallization stages causing the microstructural inhomogeneities in a cold-worked austenitic stainless steel (DIN 1.4970) after irradiation with 46 MeV nickel ions to 70 dpa. (a) Unrecrystallized, (b–c) partially recrystallized, and (d) recrystallized.

14.3 Nuclear Fuels

14.3.1 Background

The nuclear fuels of today's electricity-producing nuclear power stations are ceramics, compounds of uranium and or plutonium, most commonly the dioxides UO_2 and $(U,Pu)O_2$, the latter material usually being called MOX (mixed oxide). Most of this discussion will therefore be on UO_2 and MOX, although other fuels are or were also used. In the 1970s, so-called advanced fuels for reactors operating with fast neutrons were widely studied. These were carbides, nitrides and carbo-nitrides of uranium and plutonium. Examples of other fuels are ThO_2-based ceramics, metallic fuels, or metal–ceramic composites called cermets, coated particles—small fuel spheres coated with layers of carbon or SiC for high-temperature gas-cooled reactors, and the silicides U_3Si and U_3Si_2, mainly used in small research reactors. These latter fuels will not be treated here since they are not widely used.

The fuel materials discussed here are ceramics with high melting points, for

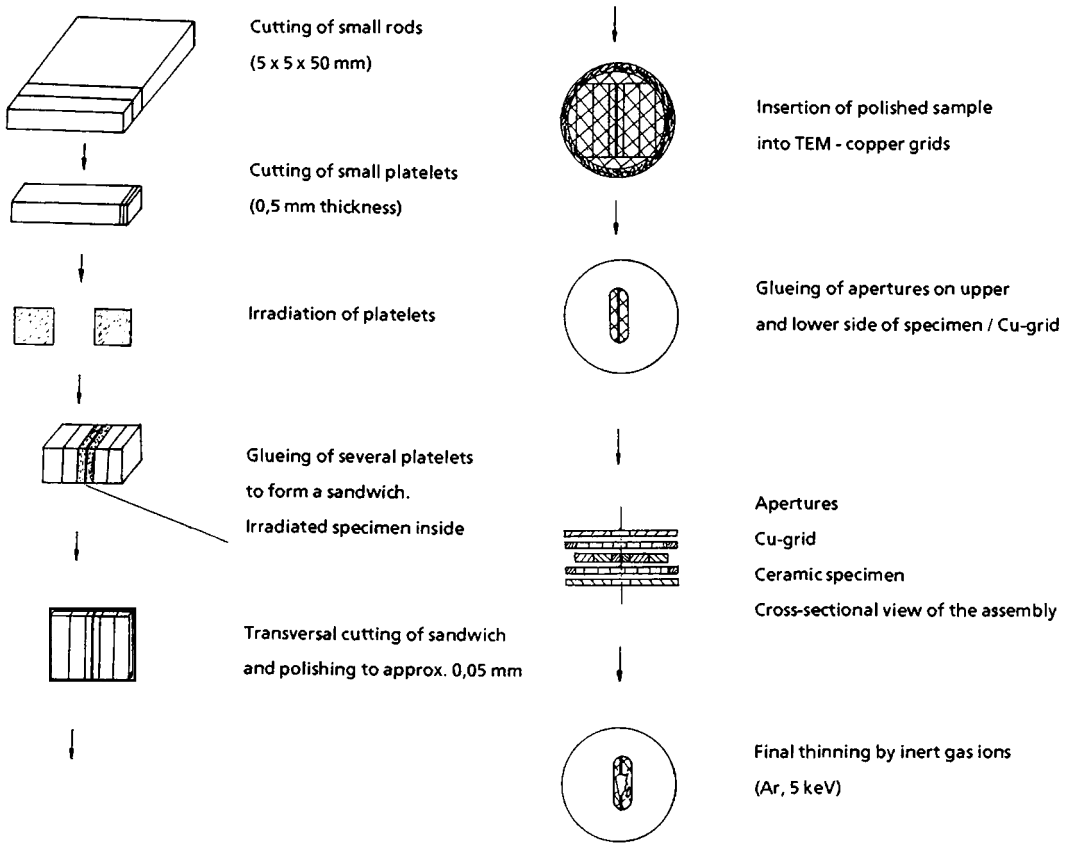


Figure 48. Cross-sectional preparation of ceramic materials for TEM investigation with a new sandwich technique.

example UO_2 (3150 K), UN (3035 K), and PuN (2843 K). The dioxides UO_2 and $(U,Pu)O_2$ have the fluorite structure, while the monocarbides and mononitrides have the rock salt structure. All these materials can be nonstoichiometric, are brittle, and have densities of $\sim 11\,g\,cm^{-3}$ (MO_2) or $\sim 14\,g\,cm^{-3}$ (MC, MN).

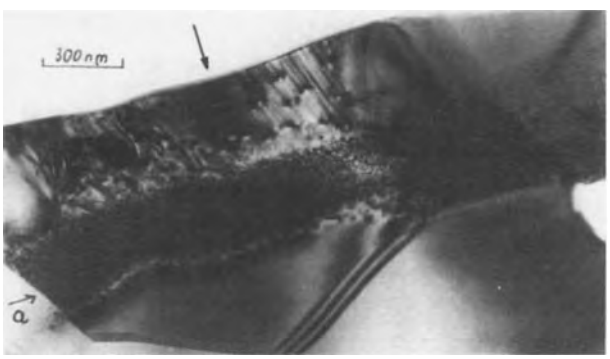


Figure 49. Cross-sectional view of SiC after helium implantation at 800°C (3×10^{21} ions m^{-2}, $E_{He} = 200$ keV). The zone of maximum radiation damage (~300 μm width) is indicated (a).

The ceramic fuel is most commonly used as sintered pellets which are stacked in metal tubes or sheaths called the clad. For water-cooled and -moderated reactors, the clad is a zirconium alloy (see Tables 14 to 18). The fuel pellets have a diameter of about 10 mm, the clad has typically a wall thickness of 0.72 mm, and the fuel rods, for LWRs, have typically a length of 4 m. The operating conditions are summarized in Table 16.

Due to the low thermal conductivity of oxide fuels, radial temperature gradients are formed during operation. Typical fuel central temperatures of LWRs are ~1200°C. Water reactors can also use heavy water for neutron economy and effective moderation of the neutrons. The most widely used type is the CANDU reactor (for *Can*adian *D*euterium *U*ranium). CANDU reactors have thicker and shorter fuel rods, and therefore the central temperatures are higher.

Reactors operating with fast (energetic) neutrons have fuels with a high plutonium content, typically 20%; the fuel rods are much thinner, typically 5.5 mm in diameter. Cooling is done with liquid metals, typically sodium. Other features are high neutron flux, high achievable burn-up, and high operating temperatures, since the high neutron energies lead to small cross-sections and enable a high power density to be obtained. Therefore, the center of the small fuel pellets gets very hot if oxides are used as fuel; temperatures above 2000°C can be reached in the center, whereas the surface of the pellet is held at about 600°C, due to the action of the coolant. Therefore, very large temperature gradients exist (up to $10^4\,K\,cm^{-1}$). If advanced fuels (MC, MN or M(C,N), where M = U + Pu) are used, these gradients are much smaller because of the much higher thermal conductivity λ of these nonoxide fuels (typical values of λ at 1000°C are $3\,W\,m^{-1}\,K^{-1}$ for MO_2 and $18–20\,W\,m^{-1}\,K^{-1}$ for MC and MN). In the fast reactor fuel, up to 20% of the metal atoms (U and Pu) can be fissioned, thus producing 40% of foreign atoms, relative to the metal sublattice. These fission products grow in as impurities at a rate of some $30\,ppm\,day^{-1}$. About 30 different elements are formed as fission products, ranging from germanium ($Z = 32$) to dysprosium ($Z = 66$)—and many of these change their chemical behavior during reactor operation due to nuclear reactions and radioactive decay. During the operating time in the reactor, radiation causes each atom in the fuel to be displaced up to and above 10^4 times. Such extreme working conditions do not exist during the operation of other, non-nuclear ceramics. A useful relation as a guideline for nuclear fuel utilization is 1 at.% burn-up of the metal atoms = 9375 MWd/tM (M = heavy metals, U + Pu). The present burn-up of LWRs is about 40 000 M Wd/t M, equivalent to the fission of ~4 at.% of uranium and plutonium.

14.3.2 Problems Treated by Optical Microscopy and Scanning Electron Microscopy

Microscopic techniques are vital to obtain oxide fuel capable of sustaining the above extreme conditions. Conventional optical microscopy and SEM are used to characterize the starting powders before pressing green pellets. The as-sintered pellets have also to be characterized with these two techniques to determine grain size and

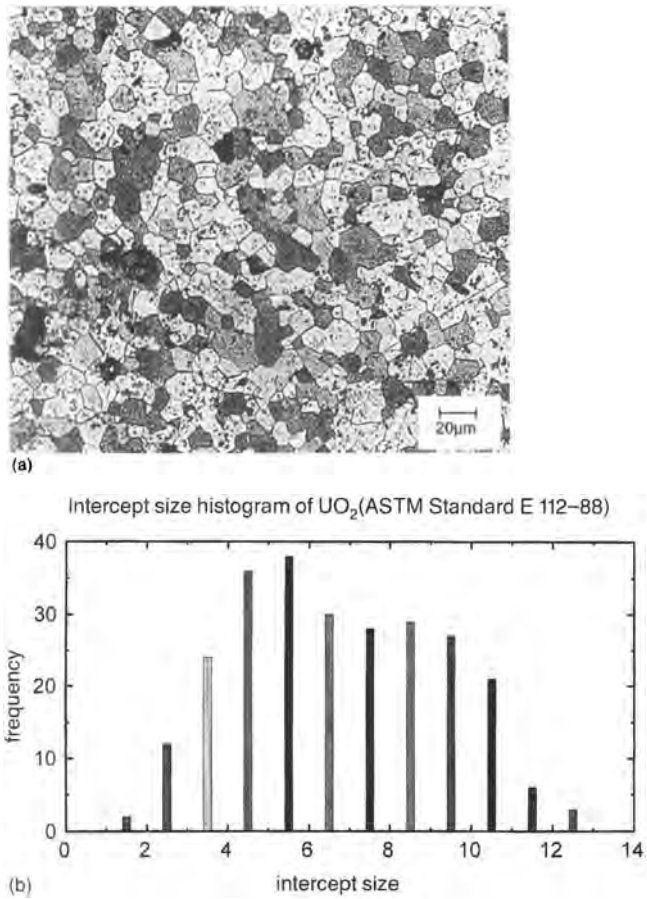


Figure 50. (a) Optical micrograph of an etched as-sintered UO_2 pellet. (b) Intercept size histogram of the same pellet (ASTM Standard E 112-88).

porosity. Figure 50 shows an optical micrograph of as-sintered UO_2, together with an intercept size histogram, applying the standard ASTM E 112-88 test method lineal intercept procedure on a 500 mm test pattern. The grain size was determined to be $9.7 \pm 1.0\,\mu m$ based on three independent intercept counts. Knowledge of the grain size is important to predict the irradiation behavior, since the grain size determines, beyond all else, the creep behavior, the release of fission rare gases, and the amount of such gases which can be accommodated at grain boundaries.

A further important result from optical microscopy and SEM is the pore size distribution. During the irradiation, the fuel matrix swells and the fission product rare gases (krypton and xenon) precipitate into bubbles. At the same time, irradiation-induced densification causes sintering pores to shrink. Figure 51 shows an unstable product. Shrinking proceeds very fast, faster than swelling, causing unwanted free volume in the fuel rod. Actually, this elimination of very fine-sized pores caused technological problems with cladding collapses in operating reactors some 20 years ago.

This densification, due to a kind of sputtering process by fission spike–pore interactions with partial escape of the

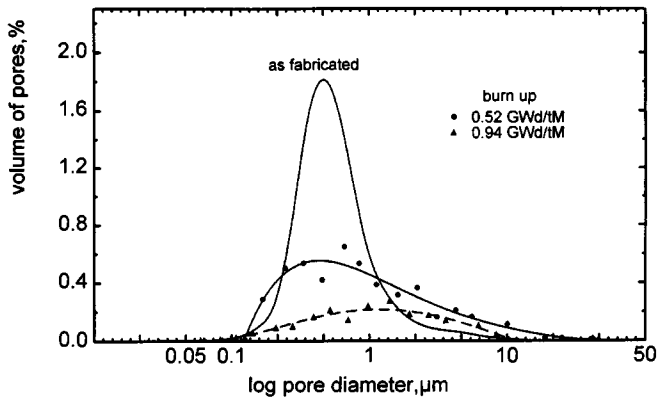


Figure 51. Effect of reactor irradiation on porosity in an unstable sintered UO_2 fuel type [113].

vacancies created in this way to grain boundaries, is less effective for bigger pores, as is obvious from Fig. 52. Two types of pellets are shown, one having a monomodal pore distribution peaked around a pore diameter of ~2.5 μm. It shows a very good dimensional stability: densification and swelling cancel one another up to a high burn-up. The second fuel with a bimodal pore size distribution shows a more pronounced swelling since the big pores do not shrink and the volume of pores usable to compensate swelling is smaller. This demonstrates the need for careful fabrication and exact microscopic investigation of the as-produced fuels.

This example of characterization and quality control shows the specific difficulty of microstructural analysis of nuclear fuels. Whereas UO_2 can be handled in controlled laboratories but without shielding, hence with normal microscope techniques, any plutonium-containing material must remain and be analyzed in a dust-tight glove box. The reactor-irradiated material has to be investigated in hot cells; see Sec. 14.1 of this Chapter. Figure 53 shows the interplay between optical microscopy in the hot cells, and the local analyses by replica electron microscopy (REM), TEM, and SEM. The gross structural changes such as cracking and pore migration in the thermal gradient to form long (columnar) grains in highly rated oxides can be deduced from optical microscopy. Finer details of the microstructure necessitate REM, TEM, and SEM techniques.

Before giving further examples of progress achieved with electron microscopy techniques on high burn-up UO_2 fuel, some very recent progress will be mentioned, obtained with high-resolution transmission electron microscopy (HRTEM) and with atomic force microscopy (AFM). Figure 54a shows the original HRTEM micrograph of a thin UO_2 specimen with lattice plane resolution. An area is shown containing a dislocation line. This is the first direct observation of a dislocation line in UO_2 obtained in February 1995 by Wang and Matzke [111] with the HRTEM instrument of the University of Albuquerque, New Mexico. Figure 54b shows the Fourier transform of Fig. 54a, Fig. 54c shows the filtered Fourier transform pattern with two of the 111 spots and Fig. 54d, finally, shows the dislocation line clearly in the inverse Fourier transform of Fig. 54c. Another

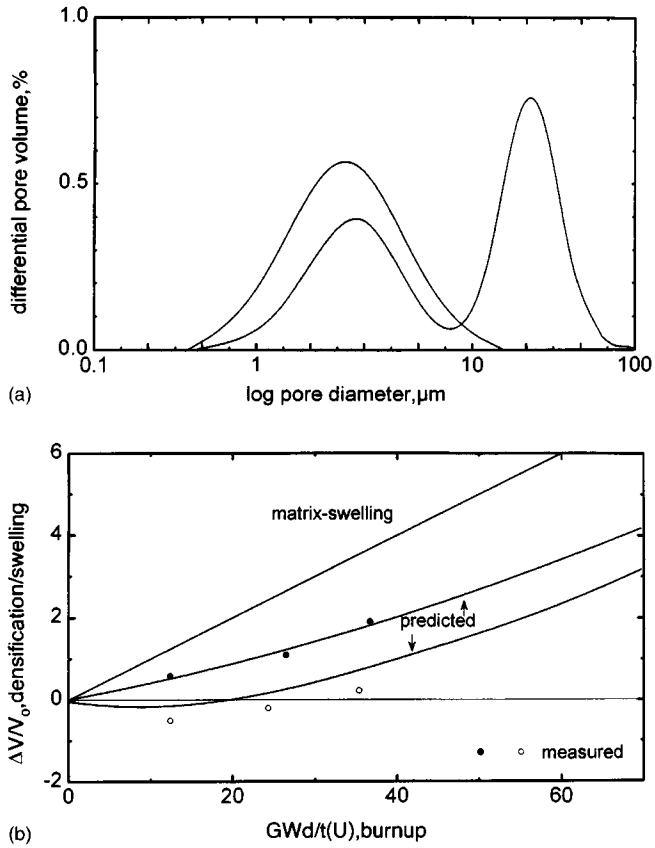


Figure 52. (a) Initial pore size distribution of two types of UO_2 pellets: 1, monomodal pores (95.7% TD); 2, bimodal pores (94.5% TD). (b) Dimensional behavior during irradiation [114].

recent advance [112] was obtained by investigating UO_2 by AFM. Small faceted voids with {111} faces could be seen in the nonirradiated material, thus giving a good 'background' to fully understand high-resolution results on the microstructure of irradiated UO_2.

14.3.3 Specimen Preparation and Electron Microscopy of Irradiated Fuel Samples

There are three techniques which have been developed for studying irradiated fuels by TEM, and these are illustrated in Fig. 53, which is a composite figure showing the combination of optical microscopy with TEM, REM and SEM on the same LWR irradiated fuel cross-section. All of the TEM methods require an initial stage of preparation within the hot cells. The methods used can be summarized as follows.

14.3.3.1 The Replica Technique

In this method a plastic film (Triafol) is used to make a replica of a polished and lightly etched fuel cross-section using remotely controlled techniques within the hot cells. The plastic replica is

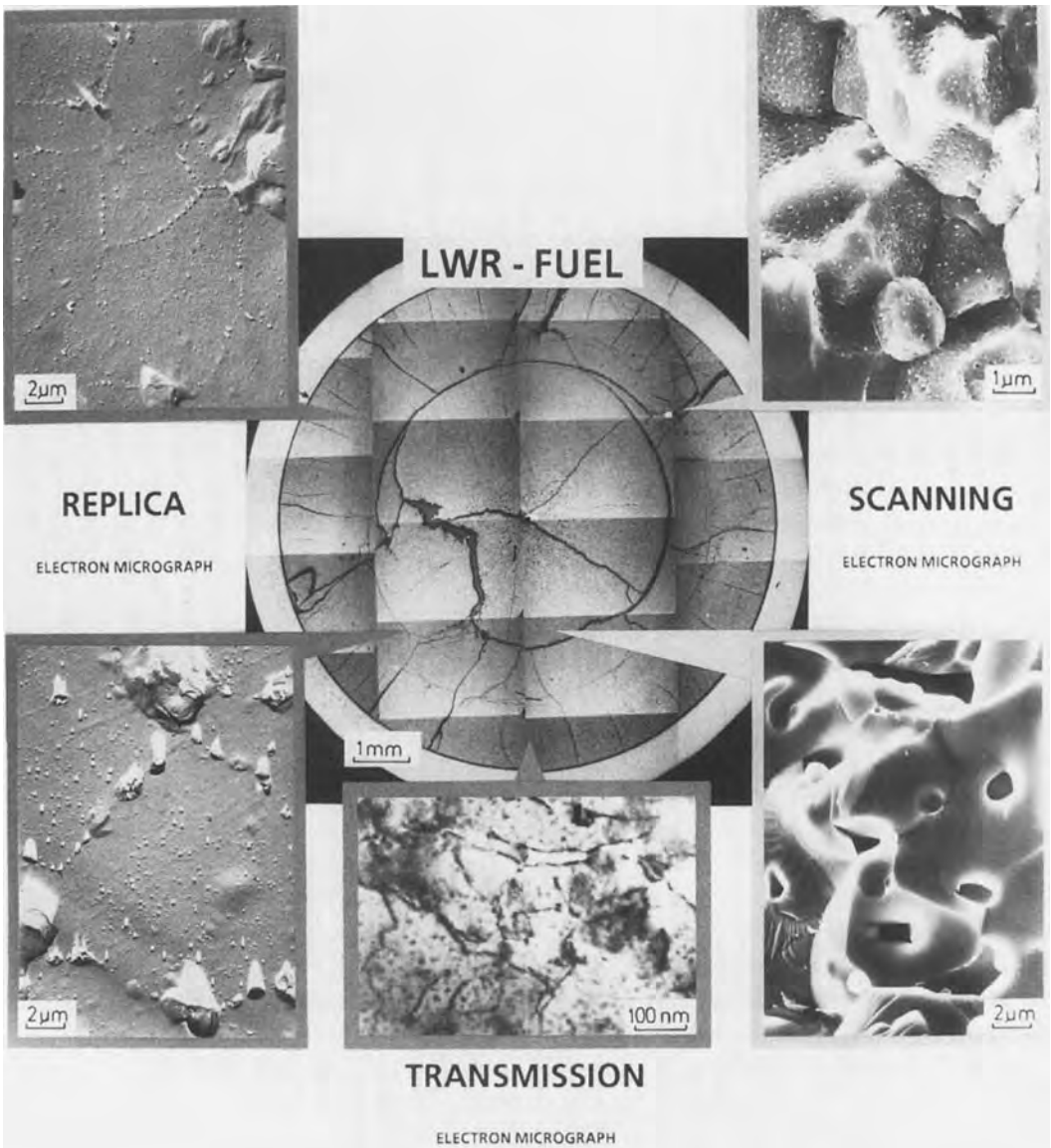


Figure 53. Cross-section of a UO_2 rod irradiated to 45 GWd/tM and subsequently transient tested at a linear rating of 41 kW m^{-1}. At the positions indicated by arrows in the central cross-section (optical microscopy), local examinations with REM, SEM and TEM were performed [115].

decontaminated using a series of ultrasonic baths, and a germanium-shadowed carbon positive replica is then prepared from this plastic film. The residual activity of the carbon replicas is usually so low that they can be handled in the open laboratory.

The information which can be obtained from a subsequent TEM analysis of the replicas encompasses inter- and

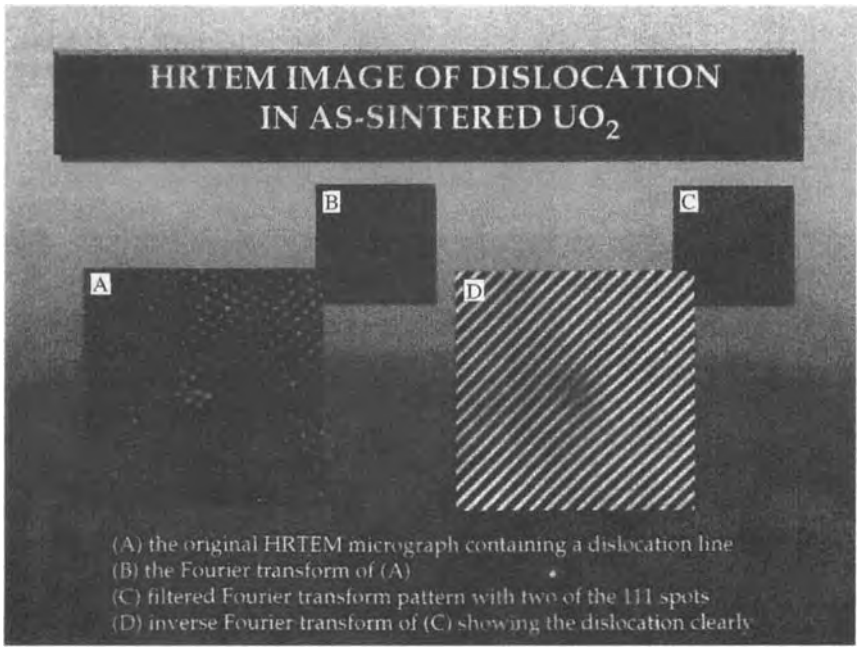


Figure 54. HRTEM of UO_2 showing lattice image resolution and a dislocation line [111].

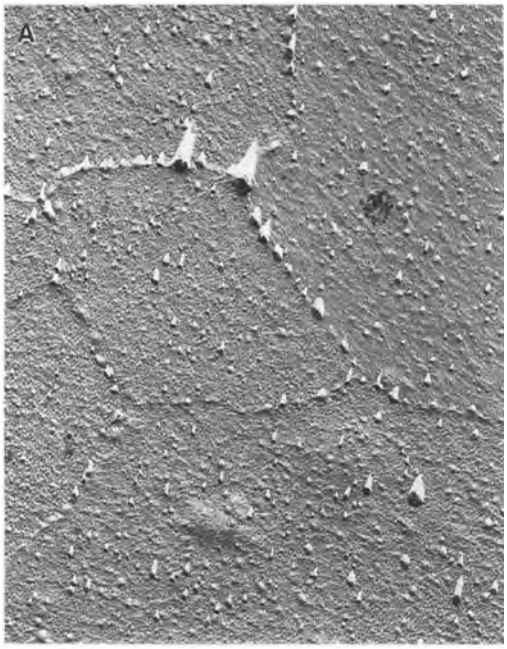
intragranular porosity and fission gas bubbles, grain boundaries, fuel cracks, and other structural features down to a resolution limit of about 30 nm. The technique has the advantage that the replicas can be decontaminated, making handling easy, and can show the structure over a very large area. Thus the full diameter of a fuel cross-section can be analyzed with a small number of replica samples, and information, for example on fission gas bubble distributions, can be obtained easily with very high radial positional accuracy. The disadvantage of the method is that features below the 30 nm resolution limit and the internal microstructure are not revealed.

The use of this method can be illustrated by the pair of replica electron micrographs in Fig. 55 taken from close to the periphery and towards the center of an irradiated LWR UO_2 fuel pin. The peripheral micrograph shows decoration of the grain boundaries by small fission gas bubbles. The central micrograph shows the development of an intragranular bubble population and high grain boundary porosity, which can be separated into grain edge porosity and grain face bubbles as demonstrated by the scanning electron micrograph of Fig. 56 of a corresponding fracture surface close to the fuel center.

Fission gas bubble size distributions and densities can be calculated from the replica images and used to construct radial fission-gas swelling profiles for the fuel cross-sections [117, 118].

14.3.3.2 Direct Transmission Microscopy of Thinned Fuel Pieces

In this method small pieces of fuel are thinned either electrolytically or by ion beam etching, and mounted directly in



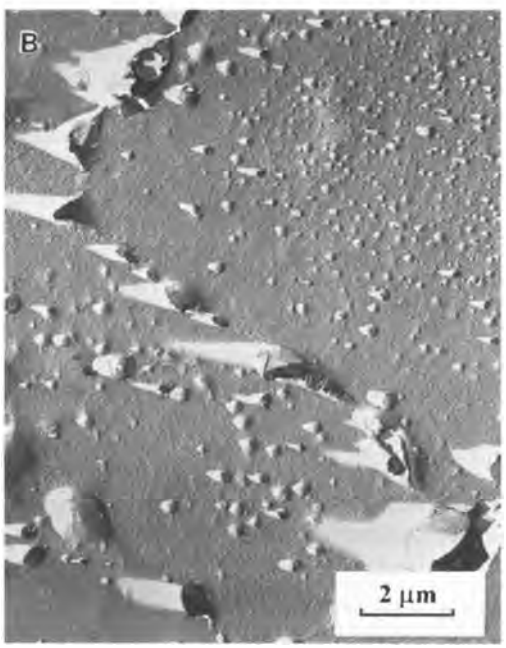


Figure 55. Replica electron micrographs taken from (A) close to the periphery and (B) towards the centre of an irradiated LWR UO_2 fuel pin. The peripheral micrograph shows decoration of the grain boundaries by small fission gas bubbles. The central micrograph shows the development of intragranular fission gas bubbles, and high grain boundary porosity.

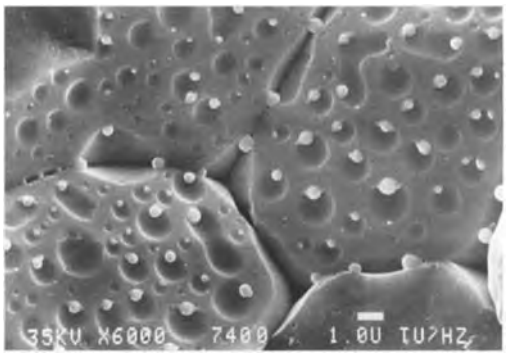


Figure 56. Scanning electron micrograph of a fractured surface of an ex-reactor LWR UO_2 fuel pellet, recorded close to the fuel centre, showing very high grain boundary porosity and a high density of spherical precipitates of the metallic fission products.

the microscope. This technique has the advantage that high-resolution analysis of the microstructure can be made, revealing, for example, very small fission gas bubbles (<10 nm), solid fission product precipitates and the dislocation or lattice defect damage structures. The disadvantage is that only very small areas can be surveyed and the positional accuracy is usually not very high.

Because the fuel cross-sections are usually heavily cracked after irradiation, it is normally not possible to cut a standard 3 mm diameter disc for the specimen preparation. A method has been developed where the fuel is separated into small irregular pieces defined by the original crack structure of the cross section, and these pieces are thinned for transmission. The pieces are typically about 2 mm in size and 0.3 mm thick. This technique has the

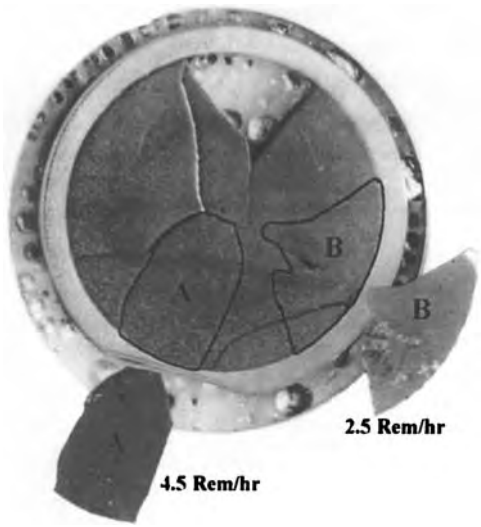


Figure 57. Small pieces of an irradiated LWR fuel cross-section which have been extracted and are of sufficiently low activity to be electropolished and examined directly by TEM. The activities of the pieces are indicated. The small pieces are about 300 μm thick, and the diameter of the fuel cross-section is about 9 mm. The curved surfaces where the pieces were in contact with the cladding permit the radial positions of the areas examined to be determined with high precision.

advantage that by referring the shape of the pieces back to optical micrographs of the full polished fuel cross section better radial positional accuracy can be obtained for the areas examined. An example of this is shown in Fig. 57. The work of separating the small pieces of fuel is performed remotely in the hot cells, but the single pieces usually have a sufficiently low β–γ activity for them to be handled with care in unshielded glove boxes.

Figure 58 shows the lead-shielded glove box which is used for the electropolishing of highly β–γ fuel samples for TEM. Biological shielding is provided by a 5 cm thick lead shield, and all operations are performed remotely using manipulators. The rear lead wall can be wheeled back if necessary to provide direct access with gloves to the box.

The right-hand compartment of the box contains a remotely driven double-jet electropolishing cell, and waste electrolyte can be drained off into a special shielded container beneath the box. The left-hand compartment of the glove box is kept clean from electrolyte and rinsing solutions, and is used for the dry handling of the samples—mounted in the Teflon holders prior to polishing and subsequent manipulations following polishing. A shielded transfer chamber attaches to this part of the cell, and this is compatible with the specimen entry port on the electron microscope glove box chain.

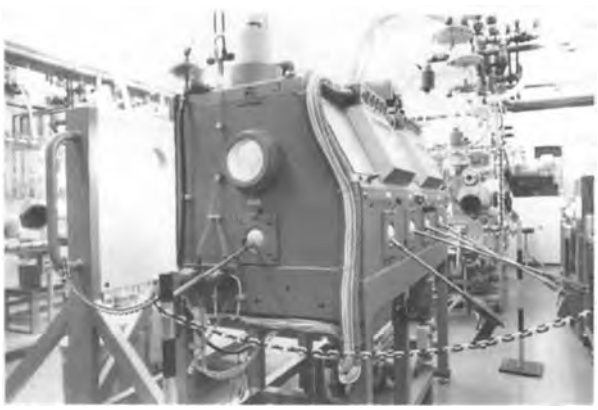


Figure 58. The lead shielded glove box used in the laboratory for the remote electropolishing of highly β–γ active samples, such as irradiated fuel.

The small fuel pieces are thinned for TEM by electrolytic polishing in a remotely operated Unithin double-jet cell mounted inside a lead cell equipped with manipulators. As the electrolyte, either Lenoir's solution or 20% phosphoric acid/80% methanol is suitable, though the latter is preferable for safety reasons inside a glove box, at a polishing potential of 20–25 V. After perforation the specimen surfaces are usually cleaned for a short time (<30 min) by argon ion etching. This preparation method is suitable for plutonium-containing oxide fuels and advanced carbide or nitrides [115, 119].

One of the main applications of TEM in the study of irradiated fuels is to investigate the behavior of the fission products as a function of the burn-up of the fuel or during a power transient, simulating a reactor incident or accident. During the steady state irradiation of an LWR UO_2 fuel, most of the fission products, both gaseous (xenon, krypton, and caesium) and solid (molybdenum, palladium, technetium, ruthenium, etc.) remain in solution in the fuel matrix, or precipitate out into very small bubbles or solid precipitates (<10 nm diameter) which can undergo fission-spike induced re-solution. Figure 59 is an example of a transmission electron micrograph of a UO_2 peripheral fuel sample with a burn-up of 4.5% FIMA, in which a very high density of small fission gas bubbles with an average diameter of 8 nm and a very narrow size distribution can be seen. The fission gas bubbles are associated with a population of precipitates of the solid fission products with approximately the same density and size distribution.

Figure 60 shows an equivalent area of fuel with similar steady-state irradiation history, but following a power transient for a period of 24 hours. Here very large fission gas bubbles and precipitates can be seen to have formed, with diameters up to 0.2 μm, leading to significant local swelling of the fuel. The precipitates are large enough to be analyzed by EDX in the microscope.

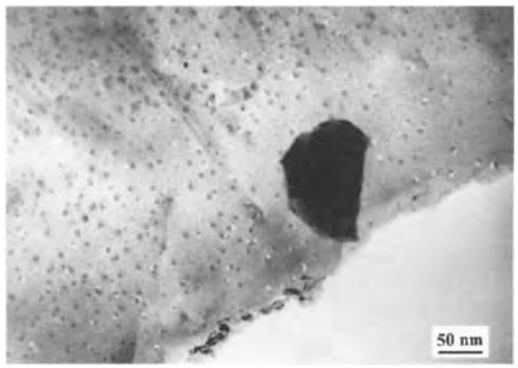


Figure 59. Transmission electron micrograph of the periphery of an irradiated UO_2 fuel sample, showing the presence of a very high density of small fission gas bubbles linked to small precipitates of the metallic fission products.

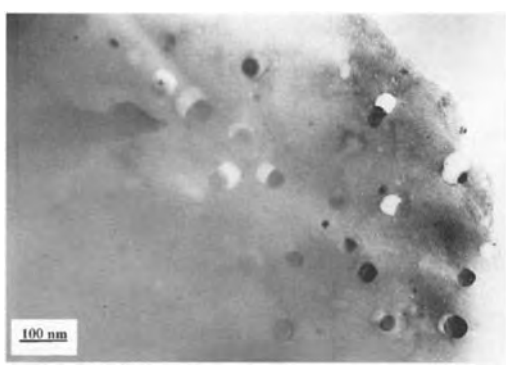


Figure 60. Transmission electron micrograph of an area of irradiated UO_2 fuel equivalent to that shown in Fig. 59, but following a power transient over a period of 24 hours.

14.3.3.3 Fuel Powder Samples

A third method of examining fuels by TEM is to use a powder sample produced either by crushing a small piece of fuel extracted from a chosen radial position on a polished cross-section, or by microdrilling into a cross-section and extracting the powder. This method has the advantage of low specimen activity, but relatively poor positional accuracy. The microdrilling must be performed within the hot cells, but small pieces can be crushed within a glove box in a standard agate pestle and mortar.

Small pieces of the fuel in the resulting powder are often thin enough for transmission. This method has the advantages of small specimen mass, thus low activity, and relatively good positional accuracy. The use of this powder method can be illustrated by TEM examinations made of the rim regions of high burn-up UO_2 fuel.

Figure 61 shows a transmission electron micrograph of a small fuel particle extracted from the rim zone of a fuel where the local burn-up is very high, in which the very small subgrain size can be seen. The inset diffraction pattern taken from the full area of the micrograph demonstrates the small difference in orientations (<5°) between the individual subgrains, which are free of dislocations or fission gas bubbles.

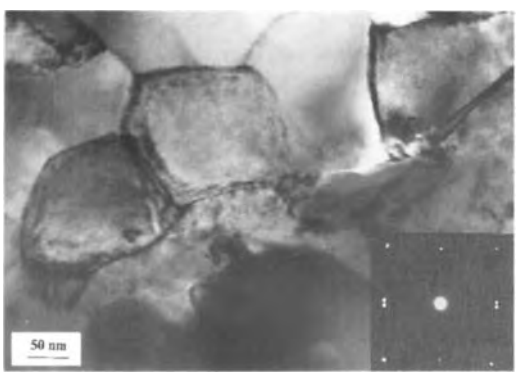


Figure 61. Transmission electron micrograph of a small fuel particle from the rim zone of a high-burn LWR UO_2 fuel pellet, showing the development of a fine subgrain structure. The inset diffraction pattern shows the small difference in orientation between the individual subgrains (<5°), [116].

14.4 Materials for Waste Disposal

For safe, long-term storage, high-level waste (HLW) originating from the reprocessing of spent nuclear fuels of electricity generating nuclear power stations is solidified in a stable matrix with high durability. Per tonne of typical reactor fuel, 300 litres of HLW are obtained in liquid form (acid solution, 2 N). This HLW is highly radioactive. The above 300 litres of HLW has a total β–γ activity of 530 000 Ci and an α activity of 2800 Ci (1 Ci = 3.7×10^{10} disintegrations per second, or 3.7×10^{10} Bq).

Today's technology of waste solidification is vitrification. Tailor-made borosilicate glasses are used, optimized for a number of properties allowing acceptability in a large-scale fabrication unit in shielded hot cells operated with remote control. These properties include viscosity of the molten glass, melting temperature, corrosion rates between molten glass and furnace and/or container, and minimization of evaporative losses during fabrication (e.g., for caesium and ruthenium). Simultaneously, acceptable glass properties have to be achieved: These include resistance against crystallization, ease of

Table 24. Composition of the (simulated) nuclear waste glass GP 98/12 (wt.%)

Material	Content (wt.%)
SiO_2	48.3
B_2O_3	10.7
Al_2O_3	2.3
Na_2O	14.9
MgO	1.8
CaO	3.5
TiO_2	4.0
Fe_2O_3	0.3
HLW[a]	14.2

[a] Contains fission products, corrosion products, and actinides (0.3 wt.% of uranium), 14.2 wt.% of HLW corresponds to ~6 mol.% because of the high molecular weight of actinide and fission product oxides. The added fission product oxides vary slightly between different glasses and are, typically: BaO, CeO_2, Cs_2O, Eu_2O_3, Gd_2O_3, La_2O_3, Nd_2O_3, MnO_2 (for Tc), MoO_2, PdO, Pr_2O_3, Rb_2O, Rh_2O_3, RuO_2, Sm_2O_3, SrO, TeO_2, Y_2O_3, and ZrO_2.

incorporating the different elements of the waste (Table 24), resistance against water attack (leaching) and against radiation. Table 24 shows, as a typical example, the composition of the German waste glass GP 98/12. This glass is a 'simulated' product, that is, the added fission products are not radioactive; the main content of the (simulated) HLW (fission product oxides, actinide oxides, and corrosion products from the reprocessing plant) are also given. Typical properties of these tailor-made, HLW-containing borosilicate glasses are: density, $\sim 2.8\,g/cm^3$; mean thermal expansion coefficient, $8 \times 10^{-6}\,K^{-1}$; thermal conductivity at 300 K, $1.1\,W\,K^{-1}\,m^{-1}$; viscosity at 1372 K, 10–20 Pa s; transformation temperature, 770–810 K; Young's modulus, $E = 83$ GPa; fracture toughness, $K_{Ic} \sim 0.95\,MPa\,m^{1/2}$ [120].

To solidify special waste streams, ceramics are studied. The basic idea originated from the fact that many natural rocks and minerals are known to accept elements of the HLW, both fission products and actinides, and to retain these elements over a long time period in a wide range of geological and geochemical environments. Typical ceramics studied as waste matrices are summarized in Table 25.

Optical microscopy, SEM, and TEM have all been extensively used to study the waste matrices. For instance, the noble metals in the HLW (palladium, ruthenium, technetium, and rhodium) form solid precipitates in the waste glass visible in the optical microscope. Or, during storage and due to accumulation of radiation damage (mainly α decay), helium bubbles can be formed, and the ceramics can be damaged to the extent that long-range periodicity of the lattice

Table 25. Some waste ceramics for actinide or HLW incorporation

SYNROC	A synthetic rock composed of (predominantly) 3–4 phases simulating minerals: perovskite $CaTiO_3$, Ba-hollandite $BaAl_2Si_2O_8$, zirconolite $CaZrTi_2O_7$ and rutile TiO_2, together with, for example, 10 wt.% of added calcined HLW
Supercalcine	Produced by adding about 50% of SiO_2, Al_2O_3, and SrO to the HLW solution. The reacted calcined product contains crystalline phases of the apatite, fluorite, scheelite, pollucite, and spinel structure
$CaZrTi_2O_7$	Zirconolite, monoclinic
$Gd_2Ti_2O_7$	Pyrochlore structure, cubic
$Ca_2Nd_8(SiO4)_6O_2$	Apatite structure, hexagonal
$ZrSiO_4$	Zircon, tetragonal
$CaTiSiO_5$	Sphene
Monazite	Mixture of orthophosphates, e.g. $(Ce,La)PO_4$

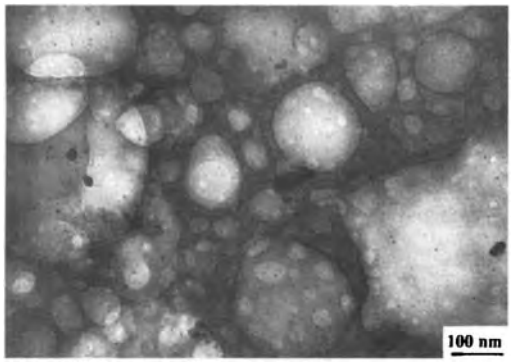


Figure 62. TEM image showing pore and bubble formation in a foil of the waste glass GP98/12 ion implanted with 4×10^{16} Xe^+ ions cm^{-2} of 225 keV energy, [121].

gets lost, that is, the materials become radiation-induced amorphous, or metamict. In glasses, radiolysis can produce free Na^+ ions and molecular oxygen, thus causing the formation of oxygen bubbles. These phenomena were studied with TEM, in the case of the ceramics in combination with electron diffraction. Figures 62–64 give typical examples.

Figure 62 shows pores/bubbles produced in the waste glass GP98/12 by ion implantation with 4×10^{16} Xe^+ ions cm^{-2} of 225 keV energy. A high density of small bubbles (5–50 nm) and a population of much larger features, resembling bubbles or pores with largely irregular shape and up to 400 nm diameter, are seen as radiation effects. The TEM results for the population of small bubbles could be shown to be in excellent agreement with parallel small-angle X-ray scattering (SAXS) results [121].

Figure 63 shows radiation damage in three potential waste matrices, $Ca_2Nd_8(SiO_4)_6O_2$, $Gd_2Ti_2O_7$, and $CaZrTi_2O_7$, containing the short-lived α emitter ^{244}Cm, to produce the α-decay

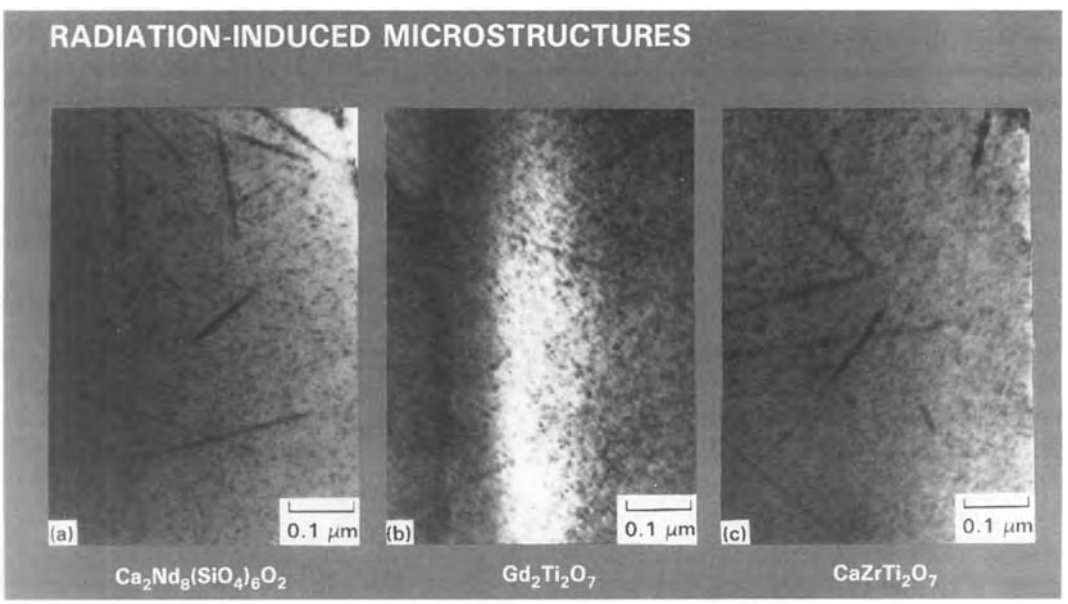


Figure 63. TEM image showing microstructural changes in three potential ceramic waste matrices of Table 25. The ceramics contained the short-lived actinide ^{244}Cm and experienced 3.6×10^{18} to 3.9×10^{18} α-decay events per cubic centimeter [122].

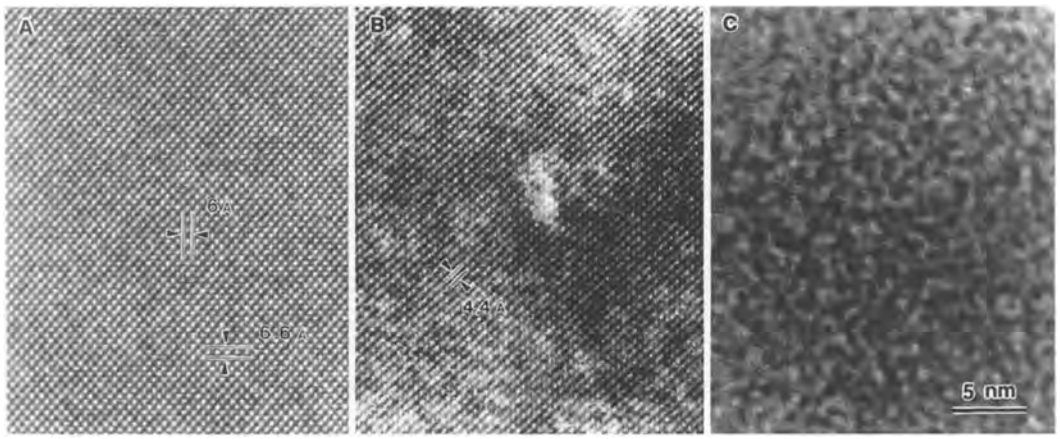


Figure 64. HRTEM image of zircon $ZrSiO_4$ before and after irradiation with 1.5 MeV Xe^+ ions. (a) Unimplanted [010] zone. (b) Same zone implanted with 1×10^{12} ions cm^{-2}. (c) Same zone implanted with the critical dose for amorphisization (2×10^{14} ions cm^{-2}). The fringes marked in (a) and (b) with 6.6, 6, and 4.4 Å show the lattice spacings for (100) (001) and (001) planes, respectively [123].

damage expected in an HLW matrix after a few 10^3 years of storage in a much shorter time period of a few years. The as-produced materials showed no specific microstructural features. Following 3.6–3.9×10^{18} α-decay events per cubic centimeter, which corresponds to 20–30% of the critical dose for amorphisization, small individual defects (~2.5–5.0 nm diameter) are seen. These defects are interpreted as amorphous, and result from the highly disordered track produced by the ^{240}Pu recoil daughter nucleus emitted with an energy of ~100 keV in the α decay of ^{244}Cm. Also seen are 'fission tracks', that is, the tracks of ~100 MeV fission products originating from the spontaneous fission of ^{244}Cm. The contribution of these tracks to swelling and amorphisization was estimated to be ~2–3% [122].

Figure 64, finally, shows damage effects in a further possible waste ceramic, that is, zircon $ZrSiO_4$. Lattice image resolution is achieved in high-resolution TEM. Damage was produced with 1.5 MeV Xe^+ ions. At small dose levels, regions with slightly mottled contrast or regions in which lattice fringes are absent are found (Fig. 64b). These regions are comparable in size to displacement cascades, indicating direct amorphization within the displacement cascades. At higher doses, complete amorphisization is observed [123].

In the above cited, and in many other publications, damage ingrowth was followed as a function of dose with TEM, HRTEM, and electron diffraction, such that a large database exists on critical doses for amorphization of different potential waste ceramics.

14.5 Electron Microscopy of Radioactive Aerosols and Particles and the Assessment of Toxicological Hazards

The release of radioactive materials into the environment, whether intentionally as a consequence of the use of nuclear weapons

or accidentally as the result of nuclear reactor accidents, represents a major area for concern. This release occurs most frequently in the form of an aerosol of small particles into the atmosphere, which can be biologically incorporated directly into the lungs, or indirectly following settlement by contact or through the food chain.

Electron microscopy plays an important role in the characterization of such aerosols and in the examination of the effects on cell tissues following incorporation. Two major areas of application can be identified:

(1) SEM is used primarily in the characterization of the aerosol particle shape and size distributions. These distributions govern how far the contamination spreads from the original source and the rate at which the particles settle.
(2) TEM is used to investigate the biological effects following incorporation of the radioactive particles.

These applications are illustrated by the following examples:

(1) The formation of radioactive aerosols under accident conditions has been investigated using simulation tests of glove box fires in which plutonium and uranium oxides have been replaced by cerium and europium oxides. The smoke produced was sampled by drawing air through filters at different distances from the fire, and particle size distributions constructed for each filter. An example of the aerosol formation is shown in the scanning electron micrograph of Fig. 65. Such micrographs also enable the interaction between soot and particles to be investigated, and permits an assessment of possible techniques for enhancing particle agglomeration [124].

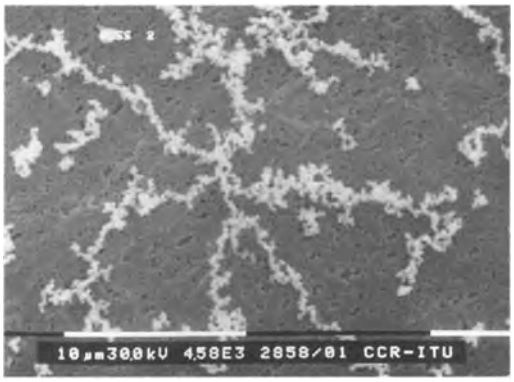


Figure 65. A scanning electron micrograph showing agglomerated soot particles on a filter, following an experimental ultrasonic precipitation test on smoke.

(2) Figure 66 shows a scanning electron micrograph of spherical $(Pu,U)O_2$

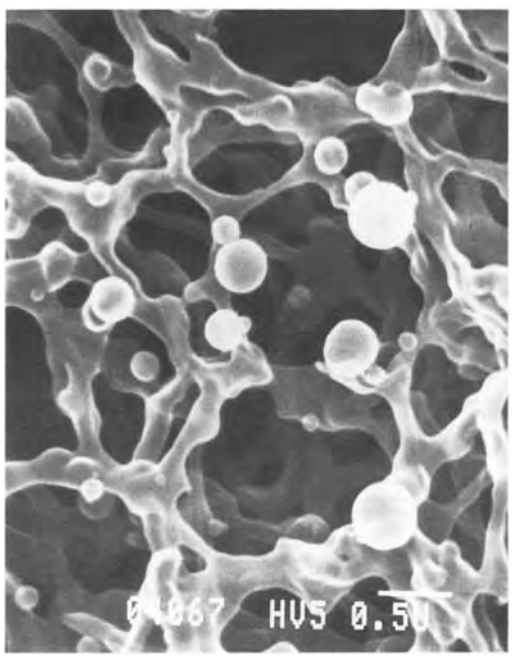


Figure 66. Scanning electron micrograph of spherical $(Pu,U)O_2$ particles deposited on a filter from aqueous suspension. With such micrographs the size distribution can be characterized prior to use in toxicological experiments on living tissues.

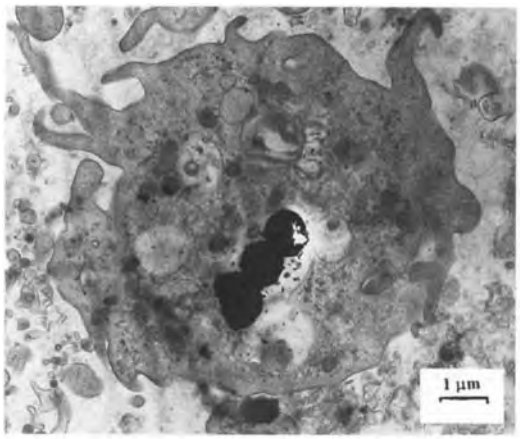


Figure 67. Transmission electron micrograph of a macrophage cell incorporating three of the $(PuU)O_2$ particles after 3 hours of incubation.

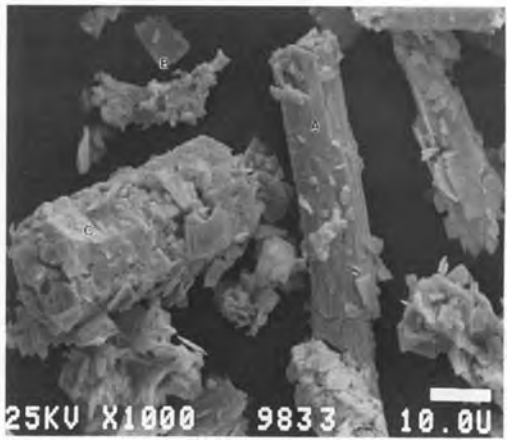


Figure 68. Scanning electron micrograph of a powder sample taken from a 'vagabond' find of radioactive materials, showing three distinct components, which could be identified by a combination of EDX analysis and electron diffraction in the transmission electron microscope. Two forms of PuO_2 were present, rod-shaped particles as indicated at A and platelets as shown at B. The third component, C, rod-like particles with a characteristic hexagonal cross-section, was identified as U_3O_8.

particles deposited on to a filter from aqueous suspension. These particles were subsequently incorporated by intratracheal injection into rats to assess the damage to the lung tissues caused by the α activity [125, 126].

The transmission electron microscope was used to follow the development of the toxicological effects of plutonium on the lung tissue. Figure 67 shows a stained section of tissue following 3 days of exposure, in which an incorporated aggregate of three of the $(Pu,U)O_2$ particles is visible. The development of the damage was studied as a function of time over a period from 1 hour to 200 days incorporation.

(3) SEM is being used to investigate the occurrence and nature of so-called 'hot spots', which are particles of very localized intense activity found in the ground following nuclear incidents such as the Chernobyl reactor accident. This is becoming an important aspect in the field of nuclear safeguards.

(4) Characterization of samples of illegal 'vagabond' radioactive materials. Very recently, electron microscopy has started to play an important role in the examination and characterization of illegal quantities of plutonium- and uranium-bearing materials which have appeared on the black market, with the objective of identifying the sources from which they have come.

The initial 'isotopic fingerprint' of the material is determined by mass spectrometry. The samples, which are usually in the form of powders, are then analyzed in the scanning electron microscope to determine whether they are homogeneous, or consist of different components. TEM is used for a structural determination using electron diffraction, and to examine the microstructure and grain size, which provide information on the original methods of production.

An example of such a 'vagabond' sample, containing a mixture of plutonium and

uranium oxides, is shown in the scanning electron micrograph of Fig. 68. This apparently homogeneous fine powder was shown in the scanning electron microscope to consist of three quite distinct components—platelets and two different types of rod-shaped particles—which could be separated and shown to have different isotopic compositions.

14.6 References

[1] *Ullmann's Encyclopedia of Industrial Chemistry*, 5th edn. (Eds.: S. Elvers, S. Hawkins, G. Schulz), VCH, Weinheim **1991,** Vol. 17A.

[2] B. R. T. Frost (Ed.) *Nuclear Materials*, Vols. 10A and 10B of *Materials Science and Engineering—A Comprehensive Treatment* (Eds.: R. W. Cahn, P. Haasen, E. J. Kramer), VCH, Weinheim **1994.**

[3] IAEA, *Safety Series No. 1: Safe Handling of Radionuclides*, 1973 edition, International Atomic Energy Agency, Vienna **1973.**

[4] J. P. Vaane in *Gmelin Handbook of Inorganic and Organometallic Chemistry*, 8th edn., Ergänzungswerk Verlag Chemie, Weinheim, Vol. 8, Transurane, Part A2 **1973,** Chap. 14, p. 320.

[5] *Safety Series*, International Atomic Energy Agency, Vienna. This covers all aspects of nuclear safety, serial numbering with No. 100 in 1996.

[6] M. Coquerelle, *Proc. of the IAEA Technical Meeting ED-WARF II, Karlsruhe, 16–20 May 1994 on Examination and Documentation Methodology for Water Reactor Fuel*, International Atomic Energy Agency, Vienna (in press).

[7] G. Wein, Karlsruhe, 1994, personal communication to E. Materna-Morris.

[8] E. Arzt, H. Fischmeister, *Pract. Met.* **1979,** *16*, 547.

[9] J. L. Chermant, M. Coster, Review on Quantitative Fractography, *J. Mater. Sci.* **1979,** *14*, 509.

[10] S. M. El-Soudani, *J. Mater.* **1990,** Oct. 20.

[11] E. E. Underwood, *J. Met.* **1986,** *38*, 30.

[12] E. E. Underwood, K. Banerji, *Metals Handbook*, 9th edn., ASM **1987.**

[13] E. Materna-Morris, *Proc. of 8th European Conf. on Electron Microscopy, EUREM, Budapest* **1984,** Vol. I, p. 795.

[14] U. Wendt, D. Wagner, Synergie von Rasterelektronenmikroskopie und konfokaler Laser-Rastermikroskopie bei Werkstoffuntersuchungen, in: *Beiträge zur elektronenmikroskopischen Direktabbildingen von Oberflächen* **1994,** pp. 27 and 113.

[15] E. Materna-Morris, O. Romer, *Proc. 18th Symp. Fusion Technology, Karlsruhe, 22–26 August* **1994,** 1281.

[16] P. M. Kelly, A. Jostson, R. G. Blake, J. G. Napier, *Phys. Status Solidi A* **1975,** *31*, 771.

[17] P. Goodman, G. Lehmpfuhl, *Acta Crystallogr.* **1967,** *22*, 14.

[18] B. L. Eyre, D. M. Maher, R. C. Perrin, *J. Phys. F* **1977,** *7*, 1359, 1371.

[19] W. Schneider, T. Stober, *Jahrestagung Kerntechnik '84* **1984,** 435.

[20] *Sicherheitstechnische Regeln des Kerntechnischen Ausschusses (KTA Nr. 3201.1)*, prepared under the responsibility of VDEh, Carl Heimanns, Köln **1991.**

[21] *Annua; Book of ASTM Standards, Iron and Steel Products*, ASTM, Vol. 01.04, Philadelphia **1987.**

[22] W. Debray, H. Cerjak, *Werkstoffeigenschaften des Stahles 22 NiMoCr 3 7 für Kernreaktorkomponenten* VGB-Werkstofftagung, Düsseldorf **1971.**

[23] H. Cerjak, W. Debray, F. Papouschek, *Eigenschaften des Stahles 20 MnMoNi 5 5 für Kernreaktor-Komponenten*, VGB-Werkstofftagung, Düsseldorf **1976.**

[24] E. Tenckhoff, M. Erve, *Materials for Nuclear Power Plants in Western Countries*, Atomwirtschaft **1992,** *37*, 185.

[25] E. Tenckhoff, M. Erve, E. Weiß, *Proc. 6th Int. Conf. on Pressure Vessel Technology, Beijing*, Pergamon Press, Oxford **1988.**

[26] C. Leitz, J. Koban, *Development of Reactor Pressure Vessel Design, Neutron Fluence Calculation and Material Specification to Minimize Irradiation Effects (ASTM STP 1011)*, American Society for Testing and Materials, Philadelphia **1989.**

[27] S. Porner, A. Seibold, R. Löhberg, *Sond. Prakt. Metallogr.* **1984,** *15*, 471.

[28] H. Cerjak, J. Schmidt, E. Breckwoldt, F. Papouschek, *Proc. 6th MPA Seminar, Stuttgart, Oct.* **1980.**

[29] J. Schmidt, E. Breckwoldt, R. Löhberg, *Proc. 9th MPA Seminar, Stuttgart, Oct.* **1983.**

[30] K.-H. Katerbau, H. Fischer, D. Kuppler. H. Ruoff, D. Blind, *Final Report on BMFT Research Project, Komponentensicherheit (Phase II), No. 1500 304 B*, MPA, Stuttgart **1989.**

[31] A. Seibold, A. Scheibe, H.-D. Assmann, *Nucl. Eng. Design* **1991,** *130*, 347.
[32] C. Leitz, *Atomkernenergie* **1977,** *29*, 75.
[33] C. Leitz, *Kerntechnik* **1987,** *51*, 256.
[34] J. T. Buswell, C. A. English, M. G. Hetherington, W. J. Phythian, G. D. W. Smith, G. M. Worrall, *An Analysis of Small Clusters Formed in Thermally Aged and Irradiated Fe–Cu and Fe–Cu–Ni Model Alloys (ASTM STP 1046)*, American Society for Testing and Materials, Philadelphia **1990**.
[35] W. J. Phythian, A. J. E. Foreman, C. A. English, J. T. Buswell, M. Hetherington, K. Roberts, S. Pizzini, *The Structure and Hardening Mechanism of Copper Precipitation in Thermally Aged or Irradiated Fe–Cu and Fe–Cu–Ni Model Alloys (ASTM STP 1125)*, American Society for Testing and Materials, Philadelphia **1992**.
[36] J. Föhl, H. Hoffmann, K.-H. Katerbau, R. Langer, H. Trunk, W. Imhof, M. Schick, R. Sinz, R. Zirn, *Langzeitverhalten, Abschlussbericht D, Forschungsvorhaben Komponentensicherheit (RS 304 A)*, MPA, Stuttgart **1983**.
[37] *Euronorm 103. Mikroskopische Ermittlung der Ferrit- und Austenitkorngrösse im Metallen*, European Community for Carbon and Steel, Beuth, Berlin, **1971**.
[38] *DIN 50 601, Metallographische Prüfverfahren. Entwicklung der Ferrit- oder Austenitkorngrösse von Stahl- und Eisenwerkstoffen*, Deutsches Institut für Normung, Beuth, Berlin, **1985**.
[39] *ASTM E-112. Standard Test Methods for Determining Average Grain Size*, American Society for Testing and Materials, Philadelphia, **1988**.
[40] E. Folkhard, *Metallurgie der Schweissung nichtrostender Stähle*, Wien, New York **1984**.
[41] A. Schaeffler, *Met. Progr.* **1949,** *56*, 680.
[42] De Long, *Wdg. J. Res. Suppl.* **1973,** *52*(7), 281.
[43] G. Homberg, G. Wellnitz, *Schweissen Schneiden* **1975,** *27*, 90.
[44] P.-H. Effertz, P. Forchhammer, J. Hickling, *Spannungsrisskorrosionsschäden an Bauteilen in Kraftwerken—Mechanismen und Beispiele*, VGB-Kraftwerkstechnik, Essen **1982,** Vol. 5, p. 390.
[45] H. Cerjak, J. Schmidt, R. Löhberg, *Proc. 4th MPA Seminar, October 1978*, Stuttgart.
[46] J. Hickling, D. Blind, *Nucl. Eng. Design* **1986,** *91*, 305.
[47] M. Erve, W. Brettschuh, N. Henzel, H. Spörl, E. Lenz, *Nucl. Eng. Design* **1986,** *96*, 217.
[48] R. L. Jones, *Corrosion* **1975,** *31*, 424.
[49] P. L. Andresen, C. L. Briant, *Conf. Environmental Degradation of Materials in Nuclear Power Systems in Water Reactors, Traverse City, Michigan* **1987**.
[50] T. U. Marston, R. L. Jones, *Proc. 5th Int. Symp. on Environmental Degradation in Nuclear Power Systems in Water Reactors, Monterey* **1991**, p. 3.
[51] M. Erve, U. Wesseling, R. Kilian, R. Hardt, G. Brümmer, V. Maier, U. Ilg, *Proc. 20th MPA Seminar, Stuttgart* **1994,** 29.1.
[52] M. Erve, G. Brümmer, H. Kleen, V. Maier, U. Ilg, H.-J. Bäumler, A. Siebold, D. Blind, *Proc. 20th MPA Seminar, Stuttgart* **1994,** 32.1.
[53] B. Stellwag, N. Wieling, L. Stieding, *Proc. 2nd Int. Symp. on Environmental Degradation of Materials in Nuclear Power Systems–Water Reactors, Monterey, California, 9–12 Sept.* **1985**.
[54] As ref. 20, Appendix 2.
[55] *Handbook of Huntington Alloys*, 5th edn., The International Nickel Company, Huntington **1970**.
[56] *Product Handbook*, Inco Alloys International Ltd., Huntington **1988**.
[57] K. Smith, *Spec. Session for Suppliers, NFA/CSN-Unipede Specialists Meeting on Steam Generators, Stockholm, Oct.* **1984**, DT/84.354, pp. 1–11.
[58] J. Orr, *Proc. of the Petten Int. Conf. on Alloy 800*, North-Holland, Amsterdam **1978,** p. 25.
[59] C. L. Briant, C. S. O'Toole, E. L. Hall, *Corrosion NACE* **1986,** *42*(1).
[60] H. Coriou, L. Grall, P. Olivier, H. Willermoz, *Proc. Conf. on Fundamental Aspects of SCC* (Eds.: R. W. Staehle, A. J. Forty, D. van Rooyen), NACE, Houston **1969,** 352.
[61] H. Coriou, L. Grall, LeGall, Vettier, *3rd Metallurgy Symp. on Corrosion, Amsterdam* **1960,** 161.
[62] H. Coriou, L. Grall, C. Mahieu, M. Pelas, *Corrosion NACE* **1966,** *22*, 280.
[63] J. L. Nelson, S. Floreen, *EPRI Final Rep. NP-3703*, Oct. **1984**.
[64] G. J. Theus, R. H. Emanuelson, *EPRI Final Rep. NP-3061*, May **1983**.
[65] Ch. Otto, U. Wesseling, *Sond. Prakt. Metallograph.* **1989,** *20*, 135.
[66] A. Mitchell, R. Brown, *Electron Beam Remelting of Zircaloys (ASTM STP 824)*, American Society for Testing and Materials, Philadelphia **1984,** p. 57.
[67] E. Tenckhoff, *Verformungsmechanismen, Textur und Anisotropie in Zirconium und Zircaloy*, Gebrüder Bornträger, Stuttgart **1980**.
[68] L. Moulin, S. Reschke, E. Tenckhoff, *Correlation between Fabrication Parameters, Microstructure, and Texture in Zircaloy Tubing (ASTM STP 824)*, American Society for Testing and Materials, Philadelphia **1984,** p. 225.

[69] D. Charquet, E. Alheritière, G. Blanc, *Cold Rolled and Annealed Textures of Zircaloy-4 Thin Strips (ASTM STP 939)*, American Society for Testing and Materials, Philadelphia **1987,** p. 663.
[70] D. Charquet, E. Alheritière, *Workshop on Second Phase Particles in Zircaloys*, Kerntechnische Gesellschaft, Bonn **1987.**
[71] D. O. Northwood, W. L. Fong, *Metallography* **1980,** *13*, 97.
[72] E. Steinberg, H.-G. Weidinger, A. Schaa, *Analytical Approaches and Experimental Verification to Describe the Influence of Cold Work and Heat Treatment on the Mechanical Properties for Zircaloy Cladding Tubes (ASTM STP 824)*, American Society for Testing and Materials, Philadelphia **1984,** p. 106.
[73] IAEA, 'Grain size determination in zirconium alloys', IAEA-TECDOC-794 **1995,** IAEA, Vienna.
[74] W. M. Mueller, I. P. Blackledge, G. G. Libowitz, 'Metal Hydrides', Academic Press, New York and London **1968.**
[75] D. L. Douglass, *The Metallurgy of Zirconium. Atomic Energy Review, Supplement*, International Atomic Energy Agency, Vienna **1971.**
[76] C. E. Ellis, *J. Nucl. Mater.* **1968,** *28*, 129.
[77] B. A. Cheadle, C. E. Coleman, J. F. R. Ambler, *Prevention of Delayed Hydride Cracking in Zirconium Alloys (ASTM STP 939)*, American Society for Testing and Materials, Philadelphia **1987,** p. 307.
[78] D. O. Northwood, U. Kasasih, *Int. Met. Rev.* **1983,** *28*, 92.
[79] O. Kubaschewsky, V. Goldbeck, *At. Energy Rev.*, IAEA Special Issue No. 6, Vienna, **1976.**
[80] P. Chemelle, D. B. Knorr, I. B. van der Sande, R. M. Pelloux, *J. Nucl. Mater.* **1983,** *113*, 58.
[81] A. Miquet, D. Charquet, *J. Nucl. Mater.* **1982,** *105*, 132.
[82] G. Maussner, E. Steinberg, E. Tenckhoff, *Nucleation and Growth of Intermetallic Precipirates in Zircaloy-2 and Zircaloc-4 and Correlation to Nodular Corrosion Behaviour (ASTM STP 939)*, American Society for Testing and Materials, Philadelphia **1984,** p. 307.
[83] G. Maussner, O. Ortlieb, H.-G. Weidinger, *Basic Properties of Zirconium Alloys with Respect to Mechanical and Corrosion Behaviour, Materials for Nuclear Reactor Core Applications*, British Nuclear Energy Society, London **1987,** p. 49.
[84] F. Garzarolli, E. Steinberg, H.-G. Weidinger, *Microstructure and Corrosion Studies for Optimized PWR and BWR Zircaloy Cladding (ASTM STP 1023)*, American Society for Testing and Materials, Philadelphia **1989,** p. 202.
[85] A. Pötzsch, A. König, A. Möckel, *Bestimmung des Ausscheidungszustandes im Gefüge de Zirconiumlegierung Zircaloy 4 und deren Bedeutung im Hinblick auf das Korrosionsverhalten*, Sonderbände der Praktischen Metallographie, München **1993.**
[86] F. Garzarolli, H. Stehle, *Proc. IAEA Int. Symp. on Improvements in Water Reactor Fuel Technology and Utilization*, Edt. IAEA, Stockholm **1986,** p. 387.
[87] D. Gilbon, C. Simenot, *Effect of Irradiation on the Microstructure of Zircaloy-4 (ASTM STP 1245)*, American Society for Testing and Materials, Philadelphia **1989,** p. 521.
[88] F. Garzarolli, P. Dewes, G. Maussner, H. H. Basso, *Effects of High Neutron Fluence on Microstructure and Growth of Zircaloy-4 (ASTM STP 1023)*, American Society for Testing and Materials, Philadelphia **1989,** p. 641.
[89] C. D. Williams, R. W. Gilbert, *Proc. Symp. on Radiation Damage in Reactor Materials*, IAEA, Vienna **1969,** Vol. 1, p. 235.
[90] D. O. Northwood, R. W. Gilbert, L. E. Bahen, P. M. Kelly, R. G. Blake, A. Jostons, P. K. Madden, D. Faulkner, W. Bell, R. B. Adamson, *J. Nucl. Mater.* **1989,** *79*, 379.
[91] M. Griffiths, *J. Nucl. Mater.* **1988,** *159*, 190.
[92] R. B. Adamson, R. P. Tucker, V. Fidleris, *High Temperature Irradiation Growth in Zircaloy (ASTM STP 1023)*, American Society for Testing and Materials, Philadelphia **1989,** p. 692.
[93] R. A. Holt, R. W. Gilbert, *J. Nucl. Mater.* **1986,** *137*, 185..
[94] A. Rogerson, *J. Nucl. Mater.* **1988,** *159*, 43.
[95] M. Griffiths, R. W. Gilbert, V. Fidleris, *Accelerated Irradiation Growth of Zirconium Alloys (ASTM STP 1023)*, American Society for Testing and Materials, Philadelphia **1989,** p. 692.
[96] W. J. S. Yang, R. P. Tucker, B. Cheng, R. B. Adamson, *J. Nucl. Mater.* **1986,** *138*, 185.
[97] M. Griffiths, R. W. Gilbert, G. J. C. Carpenter, *J. Nucl. Mater.* **1987,** *150*, 53.
[98] W. J. S. Yang, *Precipitate Stability in Zircaloy-4, NP 5591, Research Project 1250-16*, Electric Power Research Institute, Palo Alto **1988.**
[99] F. Garzarolli, W. Goll, A. Seibold, I. L. F. Ray, *Zirconium in the Nuclear Industry: 11th Int. Symp.*, Garmisch-Partenkirchen **1995,** ASTM, Philadelphia, to be published.
[100] C.-B. Cheng, R. M. Krueger, R. B. Adamson, *ASTM STP* **1994,** *1023*, 400.
[101] *Corrosion of Zr Alloys in Nuclear Power Plants. IAEA-TECDOC 684*, International Atomic Energy Agency, Vienna **1993.**

[102] H. Blank, G. Bart, H. Thiele, *J. Nucl. Mater.* **1992,** *188*, 273.
[103] H. Blank, *Z. Metallkde.* **1994,** *85*, 645.
[104] O. Gebhardt, A. Hermann, G. Bart, H. Blank, F. Garzarolli, I. L. F. Ray, *Zirconium in the Nuclear Industry: 11th Int. Symp.*, Garmisch-Partenkirchen **1995,** ASTM, Philadelphia, to be published.
[105] C. Brown, V. Levy, J. L. Seran, K. Ehrlich, R. J. C. Roger, H. J. Bergmann, *Proc. Int. Conf. on Fast Reactors and Related Fuel Cycles, Kyoto, 28 Oct.–1 Nov.* **1991**.
[106] K. Q. Bagley, E. A. Little, V. Levy, A. Alamo, K. Ehrlich, K. Anderko, A. Calza Bini, *Proc. BNES Conf. on Materials for Nuclear Reactor Core Applications, Bristol, 27–29 Oct.* **1987**.
[107] *ITER-EDA-JCT Report: General Design Requirements Document S10GDRD 195-02-03*, Febr. **1995**.
[108] K. Ehrlich, K. Anderko, *J. Nucl. Mater.* **1990,** *171*, 139.
[109] D. L. Smith, R. F. Mattas, M. C. Billone in *Nuclear Materials* (Ed.: B. R. T. Frost), VCH, Weinheim **1994,** Vol. 10B, p. 243.
[110] *Medium-term Devices and Facilities Planning in the Community Fusion Programme; Report of the European Commission, EUR 15919 EN* **1995**.
[111] Hj. Matzke, L. M. Wang, *J. Nucl. Mater.* **1996,** *231*, 155.
[112] D. Guzanos, Hj. Matzke, unpublished results.
[113] M. D. Freshley, D. W. Brite, D. L. Daniel, P. E. Hart, *J. Nucl. Mater.* **1976,** *62*, 138.
[114] G. Maier, H. Assmann, W. Dörr, *J. Nucl. Mater.* **1988,** *153*, 213.
[115] Hj. Matzke, H. Blank, M. Coquerelle, K. Lassmann, I. L. F. Ray, C. Ronchi, C. T. Walker, *J. Nucl. Mater.* **1989,** *166*, 165.
[116] I. L. F. Ray, Hj. Matzke, H. Thiele, M. Kinoshita, *J. Nucl. Mater.* (in press).
[117] C. Ronchi, I. L. F. Ray, H. Thiele, J. van de Laar, *J. Nucl. Mater.* **1978,** *74*, 193.
[118] M. Colin, M. Coquerelle, I. L. F. Ray, C. Ronchi, C. T. Walker, H. Blank, *Nucl. Technol.* **1983,** *63*, 442.
[119] I. L. F. Ray, H. Blank, *J. Nucl. Mater.* **1984,** *124*, 159.
[120] Hj. Matzke in *Ion Beam Modification of Insulators* (Eds.: P. Mazzoldi, G. Arnold), Elsevier, Amsterdam **1987,** p. 501.
[121] R. Evron, Y. Cohen, O. Regev, Y. Eyal, Hj. Matzke, K. Tinschert, *Proc. Meeting Nucl. Soc. Israel* **1994,** *18*, VIII-2.
[122] W. J. Weber, Hj. Matzke, *Radiat. Effects* **1986,** *98*, 93.
[123] L. M. Wang, R. C. Ewing, *Nucl. Instrum. Methods Phys. Res.* **1992,** *B65*, 324.
[124] L. Müller, A. Taya, G. Hotz, A. Seidel, H. Thiele, *Eur. J. Cell Biol.* **1985,** *36*, 7.
[125] H. L. Müller, E. Drosselmeyer, G. Hotz, A. Seidel, H. Thiele, S. Pickering, I. L. F. Ray, *EUR 10846 DE*, Feb. **1984**.
[126] H. L. Müller, A. Taya, E. Drosselmeyer, G. Hotz, S. Pickering, I. L. F. Ray, H. Thiele, A. Seidel, *Sci. Total Environ.* **1989,** *83*, 239.

Handbook of Microscopy
S. Amelinckx, D. van Dyck, J. van Landuyt, G. van Tendeloo

15 Magnetic Microscopy

15.1 Introduction: Observables and Methods

Below a critical temperature, the so-called Curie temperature, the elementary magnetic dipoles of magnetic materials (ferromagnetic or ferrimagnetic substances) are spontaneously ordered on a microscopic scale. This ordering can be determined, for example, by neutron diffraction techniques. Magnetic microscopy does not cover this aspect. It focuses on an intermediate, mesoscopic regime. In a volume of nanometer dimensions, many thousands of elementary dipoles are present, and micromagnetism only registers their average value, the magnetic moment density or *magnetization*. Mathematically, we have to deal with a classical vector field, the magnetization or magnetic polarization $\boldsymbol{J}(\boldsymbol{r})$.

A property of all regular magnetic materials is that the absolute value of this magnetization vector is a function of temperature alone. It is constant throughout a (chemically uniform) sample and is even largely independent of applied magnetic fields. We can therefore write $\boldsymbol{J}(\boldsymbol{r}) = J_s \boldsymbol{m}(\boldsymbol{r})$ with $\boldsymbol{m}^2 = 1$. The task of magnetic microscopy is to determine the unit vector field $\boldsymbol{m}(\boldsymbol{r})$ in the whole sample.

The magnetic microstructure $\boldsymbol{m}(\boldsymbol{r})$ is influenced by many parameters. In most materials it can easily be modified by an external field (soft magnetic materials). But it can also be pinned by lattice defects and surface imperfections, leading to hard magnetic materials and magnetic storage media. The shape of a sample is always an important parameter. In magnetic materials it is impossible to cut a sample into small pieces in order to look for the magnetic microstructure inside. The magnetization distribution would, in general, be strongly modified by the cutting process.

Unfortunately, there are virtually no microscopical techniques available to observe the interior magnetic structure of bulk samples. The methods that will be discussed in the following are either restricted to surface observations or to observations on thin films. This difficulty is partially compensated by the availability of a well-established theory called micromagnetism. The theory, which was introduced by Landau and Lifshitz in 1936 [1], is a continuum theory which derives the optimum vector field $\boldsymbol{m}(\boldsymbol{r})$ from a nonlinear and nonlocal energy functional. The external field, crystal anisotropy effects, exchange and dipolar interactions, and lattice strains contribute to the total free energy. The details cannot

be elaborated on within the scope of this contribution. It is sufficient to emphasize that due to the mentioned difficulties in experimental observation, the theoretical analysis of microscopical observations is a necessary part in successful magnetic microstructural research.

The available experimental methods in magnetic microscopy can be classified by examining the basic field equation connecting the magnetization field $\boldsymbol{J}(\boldsymbol{r})$ with the magnetic field $\boldsymbol{H}(\boldsymbol{r})$ and the magnetic flux density $\boldsymbol{B}(\boldsymbol{r})$

$$\boldsymbol{B}(\boldsymbol{r}) = \mu_0 \boldsymbol{H}(\boldsymbol{r}) + \boldsymbol{J}(\boldsymbol{r}) \tag{1}$$

which, based on the Maxwell equation $\operatorname{div} \boldsymbol{B} = 0$, leads to

$$\mu_0 \operatorname{div} \boldsymbol{H}(\boldsymbol{r}) = -\operatorname{div} \boldsymbol{J}(\boldsymbol{r}) \tag{2}$$

The magnetic field $\boldsymbol{H} = \boldsymbol{H}_{\text{ext}} + \boldsymbol{H}_{\text{d}}$ comprises any external applied field $\boldsymbol{H}_{\text{ext}}$, and also the magnetization-generated, so-called stray or demagnetizing field $\boldsymbol{H}_{\text{d}}$, which arises due to the 'charges' or poles on the right hand side of Eq. (2). The divergence of the external field is usually zero within the sample so that the field on the left hand side of Eq. (2) is just the stray field $\boldsymbol{H}_{\text{d}}$.

Figure 1 demonstrates two classical examples of magnetic microstructures, one of which does not generate a stray field, which the other does. Both possibilities are observed in nature, depending on the anisotropy properties of the material. Note that the continuous vector field $\boldsymbol{m}(\boldsymbol{r})$ of micromagnetism has been replaced in these drawings by discrete blocks or 'domains'. The formation of such domains is a well-known experimental fact. It can be justified from the mentioned theory of micromagnetism. With sufficient resolution it is possible to see that these domains

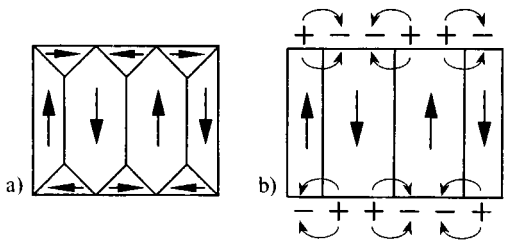


Figure 1. (a, b) Cross sections through two simplified magnetic patterns in a ferromagnetic body. For the case of the open structure (b), the magnetic charges or poles are indicated by + and − signs. The stray field extends from the positive to the negative charges, as in the corresponding electrostatic case.

are not really discrete. They are in fact connected by *Bloch walls* which form continuous transitions between the domains on the 10–1000-nm scale.

All three of the vector fields in Eq. (1) can be responsible for contrast in magnetic microscopy. Light interacts directly with the magnetization field $\boldsymbol{J}(\boldsymbol{r})$ via magnetooptical effects. Optical polarization microscopy therefore offers direct access to the magnetic microstructure. Essentially the same is true for electron polarization methods in which polarization-dependent electron scattering is analyzed, with the additional advantage of an enhanced resolution potential.

Conventional, unpolarized electrons are deflected by the $\boldsymbol{B}$-field in and around a sample. This is equivalent to a phase shift of the electron waves. Various electron microscopical imaging and scanning techniques are based on this effect.

The stray field $\boldsymbol{H}_{\text{d}}$ above a sample can be depicted by decoration with magnetic particles, leading to classical *Bitter* patterns, the method with which magnetic domains were visualized for the first time at the beginning of the 1930s [2–4]. A modern technique that is also basically sensitive to the stray field is magnetic

force microscopy, a scanning force technique using a magnetic tip. A disadvantage of the stray-field-sensitive methods is that the magnetization field determines the stray field in Eq. (2), but this relation cannot be inverted. For example, the idealized structure in Fig. 1a does not generate a stray field and hence the magnetization field cannot be inferred from any method that only scans the stray field.

There are further methods based on other effects such as the elastic deformation ('magnetostriction'), often connected with the magnetic microstructure, or effects in the inner electron shells which are somehow coupled to the magnetization vector and which can be sensed by X-ray or electron scattering methods. In practice, these new or exotic methods do not play a big role in magnetic microstructural research.

In the following we will focus on three important fields of research, which are each naturally connected with one important technique. The three fields are:

1. The investigation of magnetization *processes* in all kinds of magnetic materials. The dynamic capability and the compatibility with arbitrary applied fields of the *magnetooptic methods* make them ideally suited for this application. The available resolution (down to a few tenths of a micrometer) is fully sufficient in virtually all applications of this kind.
2. The study of *micromagnetic* details, either in thin films or at the surface of bulk materials. The enhanced resolution of various *electron microscopy* methods can be favorably used in this field.
3. Patterns in magnetic and magnetooptic *recording media* are increasingly investigated by *magnetic force microscopy*. Excellent resolution can be achieved and is needed in this application. The danger of interactions between the magnetic tip and the sample is less important because written patterns are magnetically rather hard. Also, the fact that force microscopy is primarily sensitive to the stray field becomes an advantage as the information is (at least in conventional recording) read out exactly using this field.

15.2 Magnetization Processes and Optical Microscopy

15.2.1 The Principles of Magnetooptical Imaging

The standard method of imaging magnetic domains in the optical microscope is based on a 'gyroelectric' effect which is known as the Faraday effect in transmission and as the magnetooptical Kerr effect in reflection. The electrical vector of the illuminating light interacts with the magnetization vector in a cross product fashion in the same symmetry as in the Lorentz force.

Let $\boldsymbol{E}$ be the electrical vector of an illuminating plane light wave anywhere in the sample. Then at this point a secondary amplitude

$$\boldsymbol{D}_{\mathrm{mo}} \approx Q(\boldsymbol{m} \times \boldsymbol{E}) \tag{3}$$

is generated, where Q is a (small) complex material parameter. As only transverse electromagnetic waves can exist in free space, the magnetooptic response of Eq. (3) has to be projected onto the plane perpendicular to the propagation direction of the light wave used in observation. It

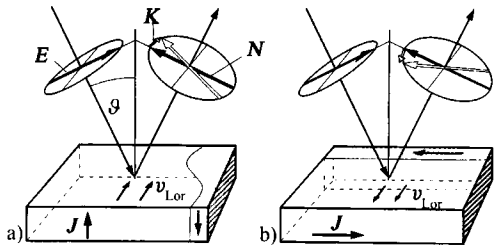


Figure 2. Schematic diagrams of (a) the polar and (b) the longitudinal Kerr effect. The interaction of the incident light ***E***-vector with the magnetization ***J*** generates the 'Lorentz velocity' v_{Lor} [parallel to the dielectric displacement $\boldsymbol{D}_{mo}$ in Eq. (3)]. This leads to a rotation of the ***E***-vector in reflection.

was shown by Wenzel et al. [5] that this simple recipe describes for small Q all possible cases: transmission, reflection, all directions of the magnetization, and even magnetic diffraction by nonuniform magnetization fields. Examples are shown in Fig. 2 for the cases of the polar and the longitudinal Kerr effect. Oblique illumination is necessary for the latter case in order to achieve a nonvanishing reaction, which consists of a small rotation of the plane of polarization of the light, changing its sign with the sign of the magnetization vector.

The magnetooptical response can be made visible in polarization microscopes as they are used in metallography or mineralogy. A high intensity light source of a limited spectral range, high quality polarizers, and the possibility to use a restricted, offset illumination aperture are important. Details of the contrast mechanism, the possible background intensities and the best modes of operation can be found in Kranz and Hubert [6]. As the effect is weak, polarization effects from imperfect surfaces and defects can easily obscure the magnetic image. The contrast can be enhanced by the application of dielectric anti-reflection coatings [6]. The most powerful option, however, is the implementation of digital difference techniques [7] which subtract a digitally stored reference image from the actual incoming image, thus getting rid of most of the disturbing background. This method is illustrated in Fig. 3. The reference image can either be a saturated image taken in a large applied field or the average of moving domain pictures taken in alternating fields of sufficient amplitude.

It is instructive to take a look at the contrast mechanism for the last-mentioned case of a digital difference image. Let φm_x be the (real part of the) magnetooptic rotation angle caused by the magnetization component m_x, and let β be the analyzer angle counted from the position of extinction (Fig. 4). Then the intensity passing through the analyzer is $\sim \sin^2(\beta - \varphi m_x)$. Assuming the reference

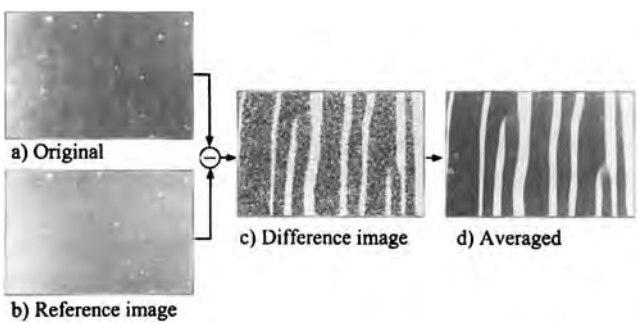


Figure 3. Illustration of the digital difference technique. In (a) the original, the domains on a Permalloy film are barely visible. Digital subtracting (b) the saturated reference image leads to (c) the difference image, which, after averaging several frames, results in (d) a clear and low noise picture.

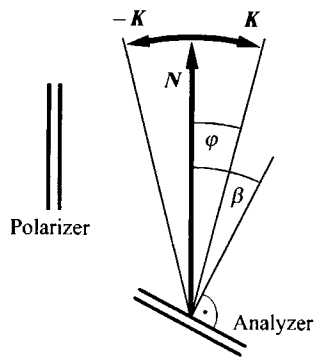


Figure 4. The principle of magnetooptic contrast formation. The regularly reflected amplitude N of the polarized light is modified by a small magnetization-dependent magnetooptical amplitude K, leading to rotation of the polarization plane by a (small) angle φ. This rotation is made visible by an analyzer which is offset by a larger angle β. If K differs from N in phase, the rotation is connected with some ellipticity. In this case the phase should first be adjusted by an additional compensator.

image is formed at $m_x = 0$, we obtain for the difference signal

$$S = I_0 + I_1[\sin^2(\beta - \varphi m_x) - \sin^2(\beta)]$$
$$\approx I_0 + 2I_1\beta\varphi m_x \qquad (4)$$

where I_0 is a chosen bias level and I_1 is the regularly reflected light intensity. The approximation is based on φ, $\beta \ll 1$ and $\varphi \ll \beta$.

We note a number of important properties:

1. The signal is *linear* in the magnetization component m_x observed, a property not often found in microscopy.
2. The intensity of the signal can be tuned by adjusting the analyzer angle β over a wide range to adjust to the sensitivity of the detector.
3. Assuming shot noise as the basic noise mechanism, the noise should be proportional to the square root of $I_1 \sin^2(\beta) \approx I_1\beta^2$. The signal-to-noise ratio is then proportional to $K = \varphi\sqrt{I_1}$, which is the magnetooptical amplitude shown in Fig. 4. The noise content of an image is therefore not influenced by changing the analyzer angle β, as long as it stays large compared to the magnetooptical rotation φ and small compared to unity.
4. If the magnetooptic signal K contains an imaginary part relative to the regularly reflected light amplitude N, an adjustable phase compensator should be inserted to convert the magnetooptical signal into a detectable rotation.

With different microscope settings, different magnetization components can be observed separately. This is best achieved by adjusting the illumination diaphragm inside the full aperture of the microscope. A central aperture, corresponding to perpendicular incidence, only detects the polar magnetization component. Choosing two different off-center positions of the diaphragm, corresponding to two orthogonal planes of incidence, both in-plane components of the magnetization vector can be detected. Such 'quantitative' procedures are discussed in the literature [8–10]. Possible interference between different effects should always be considered. Reliable results require some kind of contrast calibration.

The magnetooptic Kerr effect is not sensitive to the reflecting surface alone. The signal is formed by linear superposition of contributions from various depths, weighed in phase and sign by the amplitude penetration function of the incident and the secondary emitted light, as explored in detail in Traeger et al. [11]. Choosing different phase conditions, the depth sensitivity characteristics can even

be adapted according to the specific aims of the experiment [12].

15.2.2 Magnetization Processes in Thin Film Elements

Many applications such as magnetic sensor or memory devices are based on magnetic thin film elements. Their magnetization process can be quite complex, but still less involved than the processes in extended, bulk magnetic materials in which far-reaching and local interactions act together, forming complex three-dimensional domain structures. Thin film elements do not hide much in the third, i.e., the thickness dimension, and the whole sample can be observed at the same time. The behavior of such elements is demonstrated for two examples.

The first thin film element has a circular shape and actually consists of many exchange-coupled sublayers which act, however, as a homogeneous, nearly isotropic, soft magnetic film. The demagnetized state in such circles can consist of a simple concentric pattern (Fig. 5a), but also of much more complex states depending on the history of the applied field. Figure 5b–f demonstrates how the simple pattern in (a) is converted into a concentric two-domain pattern (f) via a sequence of applied fields which lead to the formation of a domain wall (b) and the interaction of this wall with the sample edge (c–d). More details on these samples can be found in Hubert and Rührig [13].

In the second example the magnetization process of an elongate rectangular Permalloy ($Fe_{81}Ni_{19}$) element is studied. Such elements display a near-saturation high-remanence state if their aspect ratio is larger than about 6:1 [14]. This does not mean, however, that they are completely saturated. Both ends always carry a kind of closure pattern which avoids strong magnetic charges and replaces them by a more distributed charge pattern. A quantitative evaluation of such a closure pattern is shown in Fig. 6.

On applying an external field opposite to the remanent magnetization, the closure patterns expand from both ends until they meet in the middle. If the two closure configurations happen to be oppositely magnetized in their transverse components,

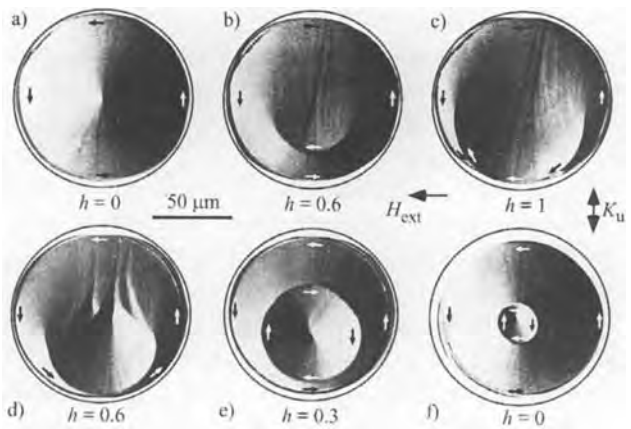


Figure 5. Magnetization process in a soft, magnetic, circular disk. (a) The simple concentric state spontaneously forms (b) a domain wall in an applied field, which, when touching the sample edge (c), leads to (d, e) an irreversible domain rearrangement ending in (f) a metastable concentric two-domain state. The external field is given by h, the reduced field relative to the theoretically derived saturation field for the inner part of the disk.

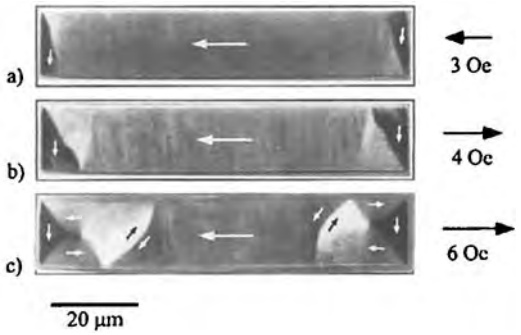


Figure 6. Closure patterns in a rectangular element in a high-remanence state. The sample consists of a single layer of Permalloy of 17-nm thickness. Depending on the strength and sign of the longitudinal field, the domains at the sample ends shrink and expand, while the central part of the element stays saturated. (1 Oe $\approx$ 80 A m^{-1}.)

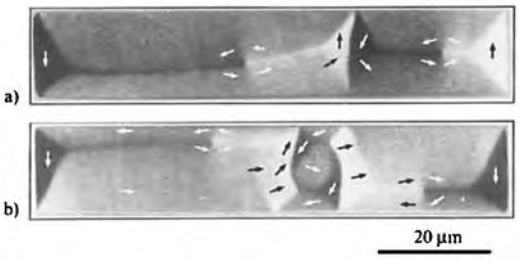


Figure 7. Zero-field result of a collision of expanded closure patterns: two cases. The variant (a) develops when the edge closure domains are magnetized in opposite directions. The complex pattern (b) is only found as a metastable structure when the two end domains are oriented along the same direction. A quantitative analysis leads to the inserted magnetization direction arrows. Sample as in Fig. 6.

they can form a domain pattern after their collision which is related to the well-known cross tie pattern and which is shown in Fig. 7a. Alternatively (Fig. 7b), an awkward metastable residual pattern is formed in the middle [15] which can barely be understood other than by quantitative experimental evaluation.

In a field parallel to the remanent magnetization in the center part, the closure pattern shrinks, but it remains stable up to remarkably high field values, as shown in Fig. 8. The complete disappearance of the closure pattern cannot be directly observed with optical means. It can be derived from experiments, however. If a high field does not destroy the pattern completely, the residual domains act as

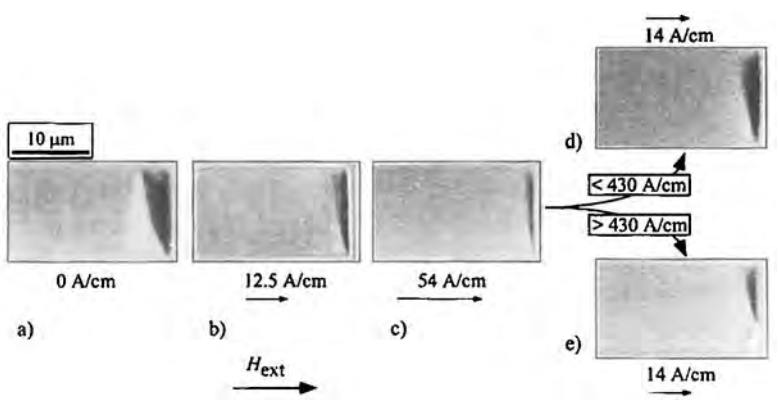


Figure 8. High field behavior of closure patterns in rectangular elements. The last residual domains in higher fields than in (c) cannot be seen with optical means. Their presence can only be inferred from the behavior of the domains when the field is reduced again. Up to a field of 430 A cm^{-1} a pattern as in (d) is observed which corresponds to (b). Beyond this threshold a new symmetric pattern as in (e) is formed, which is only converted back into the lower energy asymmetric configuration at negative fields. Sample as in Fig. 6.

nuclei, and closure patterns of the same type will reappear after reducing the field. If, on the other hand, the field was high enough to annihilate the closure structure, a newly formed, almost symmetrical closure pattern forms. Experiments based on this criterion revealed that residual domains persist in fields that are several times as large as needed to make the domains invisible. This behavior of residual domains is quite typical in ferromagnets and leads to important consequences in hard magnetic materials. To achieve the highest coercivity, many such materials have to be magnetized first in very high pulse fields of many thousand kiloamps per meter.

15.2.3 Domains and Magnetization Processes in Bulk Crystals

Oxidic magnetic materials such as ferrites or garnets are transparent to visible light or at least to infrared radiation. In these materials the internal magnetic structure can in principle be investigated. A good example is given in Fig. 9 [16]. Here the highly complex surface structure made visible by the Bitter pattern technique is shown to be formed on top of a smooth internal structure which becomes visible by infrared transmission microscopy. This mechanism is called domain *branching*; it occurs if the easy direction in the bulk of a sample is incompatible with the surface orientation and tries to reconcile a low energy, simple domain pattern in the volume with a finely divided surface pattern [17].

An interesting feature of these branched domain patterns in uniaxial crystals led to a widely accepted application. On increasing the thickness of such a crystal, the internal complexity increases continuously, leading to wider and wider internal domains. The surface domain width, however, stays constant [18], independent of the crystal thickness. This is determined by an equilibrium between the wall energy and the demagnetizing energy. From its measurement the wall energy can be conveniently determined [19].

Metallic materials do not offer the opportunity to look inside or through bulk crystals (except with expensive and at the same time quite limited techniques

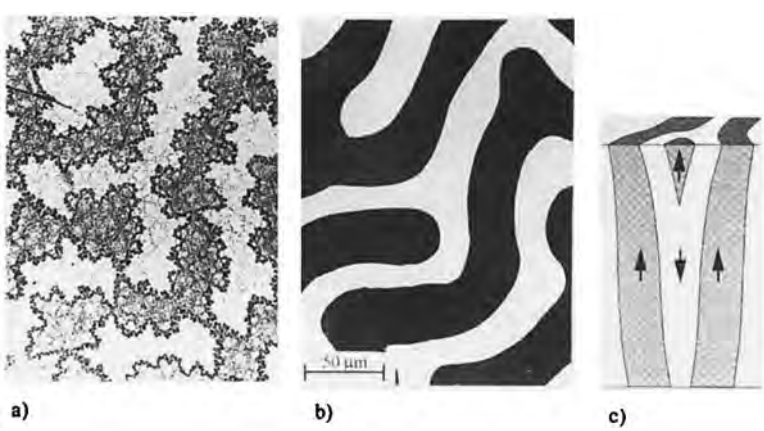


Figure 9. Domains of a hexaferrite crystal plate (a) on the surface and (b) in infrared transmission. The qualitative interpretation (c) indicates how wide internal domains can coexist with fine surface domains. (Courtesy of R. Szymczak, Warsaw.)

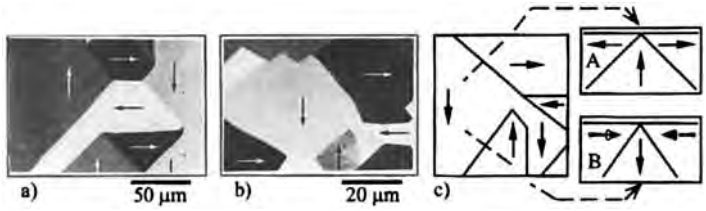


Figure 10. Domains on a (100)-oriented silicon iron crystal. The simple domains in (a) are found on stress-free crystals. Slight mechanical stresses may favor the perpendicular easy direction and lead to the appearance of V-lines as in (b). Their internal structure is indicated in (c) by a schematic top view and two cross sections.

such as X-ray topography [20, 21] and neutron topography [22, 23]). Some theoretical 'insight' is necessary to understand what is going on under the easily observable surface. Examples of observations and interpretations are shown in Fig. 10.

The possible domain directions in such crystals are determined by the preferred crystal directions which are the ⟨100⟩ directions in iron. Most of the observed domain walls offer a natural interpretation of two such domains separated in a stray-field-free manner. Such walls can immediately be recognized by their well-defined and straight orientation (Fig. 10a). Different-looking domain walls appear if, induced for example by mechanical stress, the additional easy directions perpendicular to the surface are favored. The 'domain walls' visible under such conditions are actually intersection lines of two internal domain walls which meet at the surface. These lines are called V-lines. They are explained in the cross sections of Fig. 10c.

External fields can also lead to the formation of internal domains and to a complete rearrangement of the domain structure if the field is applied along a hard direction. This is illustrated in Fig. 11 for a transformer steel material, i.e., a (110)-oriented silicon iron crystal. A field is applied parallel to the [111] direction rather than along the easy [001] axis. The complex domain patterns formed in this magnetization process demonstrate the extra magnetization losses that occur if a transformer material is activated along a disfavored axis. In the highest fields,

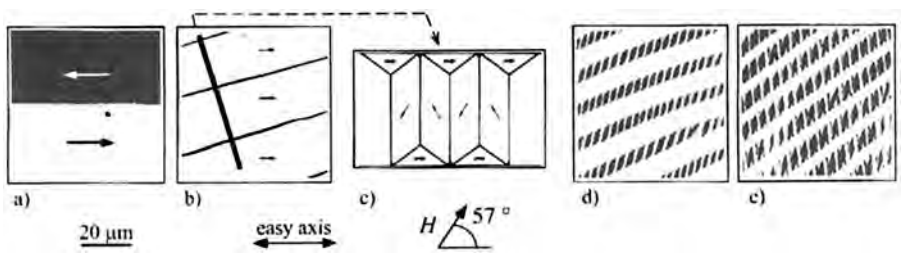


Figure 11. Domain rearrangement processes on a (110)-oriented SiFe transformer steel crystal in increasing applied fields along the hard direction [111]. The simple, field-free state (a) is only shown for reference. The model in (c) shows a cross section through the first field-induced pattern (b) which occurs at about two thirds of saturation. The internal structures of the later, heavily branched structures (d, e) have never been explored in detail.

complex surface patterns are formed which are related to the branched patterns of uniaxial crystals (see Fig. 9).

15.3 Micromagnetics and High Resolution Electron Methods

15.3.1 Optical Investigation of Micromagnetic Details

Exploring the details of domain walls and their substructures contributes to the fundamental understanding of magnetic materials. Important technical properties are directly linked to domain wall processes. The high permeability of soft magnetic materials, which can easily exceed 10^5 times the permeability of air, is a consequence of the ease of displacing domain walls in undisturbed crystal lattices. On the other hand, modern hard magnetic materials rely on the high energy and narrow width of their domain walls. (A high wall energy makes nucleation of new domains difficult, a narrow wall width favors an effective pinning of walls by lattice defects.)

The fastest and most economical access to magnetic microstructures is always optical microscopy, if it is applicable. The introduction of digital image processing opened this field to optical methods which had been unsuitable except for transparent oxide crystals. Figure 12 shows the beautiful substructures of domain walls visible without any special techniques in garnet crystals. No resolution problem appears to exist in the case of these very wide walls. The characteristics of the various wall types agree with theoretical predictions based on the anisotropy and magnetostriction material data.

Substructures of opposite wall chirality, as visible in Fig. 12, can only be indirectly inferred in metallic samples. However, a special asymmetric surface structure of domain walls becomes visible in soft magnetic materials. The polar magnetization direction in the interior gives way to a surface vortex which closes most of the magnetic flux. Observations and their interpretation are shown in Fig. 13. The slight tilting of some wall segments is a

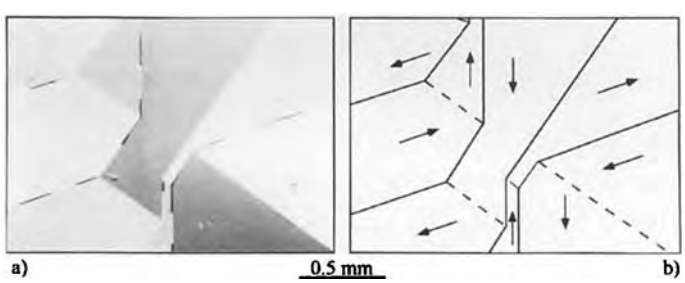


Figure 12. Domains and walls on a (110)-oriented garnet crystal observed by the Faraday effect. Mostly the domain walls contribute to the contrast. Note that both 180° walls and 109° walls are widened and often subdivided into segments of opposite polarity, while 71° walls [dashed in (b)] appear narrow, uniform, and somewhat diffuse. The domains are also visible due to a slightly nonperpendicular illumination. The domains structure is explained in (b), based on the ⟨111⟩ easy directions of the magnetic garnet.

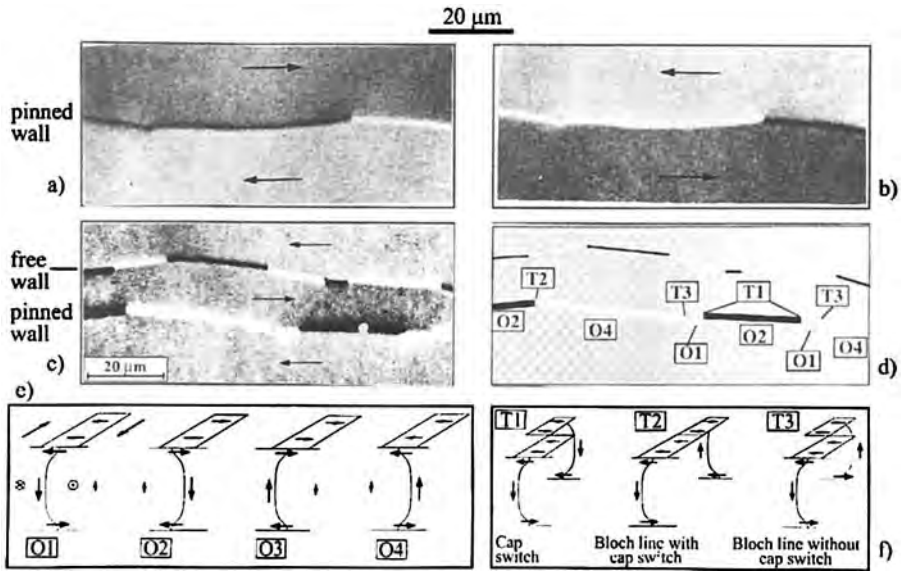


Figure 13. Different wall states compatible with a baked-in wall pattern in a nearly magnetostriction-free metallic glass of 30-µm thickness. Variants of the pinned wall, which are both compatible with the baked-in anisotropy pattern, are shown in (a, b). An additional 'free wall' with an arbitrary internal structure appears in (c). The internal structure of the pinned wall in (c) is indicated in (d–f), listing all possible wall orientations and wall transitions.

consequence of the vertical flux, indicating domain walls of opposite chirality. Some of the transitions are not connected with a kink. They represent mere 'cap switches', a change in the orientation of the surface vortex without a change in the basic wall chirality.

The observations in Fig. 13 were carried out on a metallic glass which had been annealed in the presence of a domain wall. Along the line at which the wall has been 'baked in', a special induced anisotropy pattern is left. A 'pinned' wall that occupies this position is particularly wide. The interaction of this pinned wall with 'free' walls leads to particular magnetization loops which can be used in special sensor applications [24].

The peculiar surface structure of domain walls in soft magnetic materials was first predicted by theory [25]. The first experimental confirmation was achieved by a magnetooptical scanning technique [26]. Optical microscopy was tried for a long time without success (partly also because researchers looked for the absent polar component at the surface), until the digital difference image technique offered an easy access to this world [7]. Optical microscopy can even see domain walls that are narrower than the resolution limit, although the image of these narrow objects is, of course, widened. Integrating over the image gives access to the true domain wall width. This is a consequence of the linear relation [Eq. (4)] between image and source in magnetooptical imaging, and was demonstrated, for example, by Schäfer et al. [27]. Nevertheless, there is no way of measuring the detailed profiles of

domain walls with optical methods, let alone the structure of their subunits such as the so-called Bloch lines, i.e., the transition between wall segments. Higher resolution methods are available in electron microscopy, as will be discussed in the following sections.

15.3.2 Electron Polarization Techniques in Scanning Electron Microscopy

Secondary electrons emitted from a clean-surface ferromagnet are spin-polarized [28], the sign of the polarization being opposite to the magnetization vector in the surface of the material. Scanning a sample with electrons and analyzing the secondary electrons offers a direct, high resolution, and quantitative image of the surface magnetization of the sample. This idea was first demonstrated by Koike and Hayakawa [29] and soon after by other groups [30–32]. The acronym SEMPA (secondary electron microscopy and polarization analysis), introduced by Unguris et al. [30], has been widely adopted for this technique. Figure 14 shows an example of a quantitative, high resolution image obtained in this way [33]. While Kerr microscopy is limited to about 200 nm, a resolution of better than 20 nm was demonstrated for the electron polarization technique [34]. Every electron polarization picture displays one component of the magnetization vector. Two components can be recorded simultaneously with two sets of detectors. The third component is accessible after an adequate spin-rotation operation.

Scanning the recorded images immediately gives the profiles of the chosen magnetization component through the walls for the top one or two atomic layers of the sample. In Scheinfein et al. [35] a convincing agreement between numerical micromagnetic calculations and measured surface profiles was demonstrated for an iron crystal. Polarized electron techniques are also particularly valuable for the investigation of ultrathin films, as demonstrated in Fig. 14 [36].

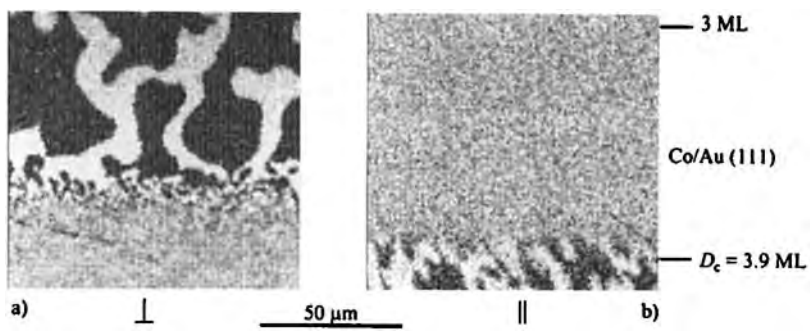


Figure 14. High resolution image of a magnetic microstructure in an ultrathin, 'wedge-shaped' cobalt film, the thickness of which increases from top to bottom. The observed pattern indicates a transition from a preference to the vertical direction for films that are thinner than about 0.6 nm (top) to a planar structure for thicker films (bottom). The two different magnetization components [(a) vertical and (b) parallel to the thickness gradient] are made visible by electron polarization analysis in different detectors. Interesting is the finely divided domain structure in the transition region. (Courtesy of S. Speckmann, H. P. Oepen, Jülich and Halle.)

The limitation of the electron polarization method still lies in its unfavorable imaging statistics. First, the used electrons have about a thousandfold higher energy than the photons used in optical microscopy. For a given sample energy load, the number of quanta per time interval is necessarily much smaller, leading to correspondingly larger exposure times for the same image quality. This applies to most nonoptical microscopical techniques and constitutes a disadvantage of these methods even if the relative strength of the basic physical effects may be larger than in the case of the magnetooptical effects.

In the case of electron polarization analysis, the problem is aggravated by the low efficiency of available polarization analyzers. While optical analyzers lose none (for crystal analyzers) or at most one half (for sheet analyzers) of the favorably polarized photons, electron analyzers have an efficiency of the order of 10^{-4}. Much effort has been spent in order to push this value into more favorable ranges. Until this goal has been achieved, electron polarization microscopy will have its place among specialized investigations of details in the surface magnetization of bulk materials which cannot be performed by other methods.

15.3.3 Quantitative Transmission Electron Microscopy Techniques

Transmission electron microscopy of magnetic materials is based on the Lorentz deflection of the electrons and is therefore independent of electron polarization. The optical analogue of a magnetic specimen consisting of several domains is a set of flat prisms, as shown in Fig. 15. The prism analogy also demonstrates that magnetic contrast is a kind of phase contrast: The different thicknesses of the glass generate corresponding phase differences of the light waves. Any of the different options of phase microscopy can be applied in electron microscopy in order to make magnetic information visible. An excellent introduction to these techniques entitled 'Lorentz Microscopy' can be found in Chap. 4.1.7 of Vol. I of this Handbook.

Defocusing is the oldest such technique [37], also called the Fresnel method from the Fresnel interference experiment. It is

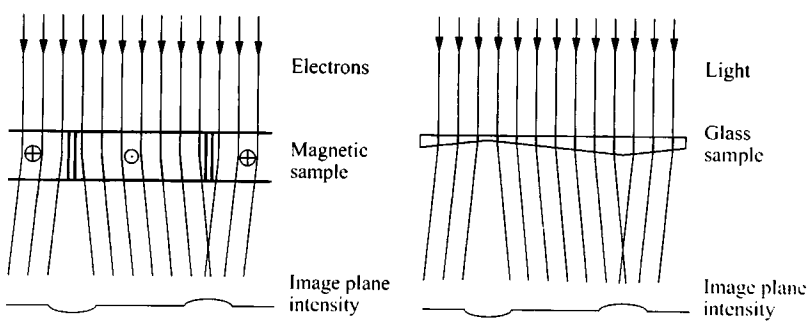


Figure 15. The analogy between magnetic electron deflection and optical deflection by prisms, at the same time illustrating image formation in the defocused mode.

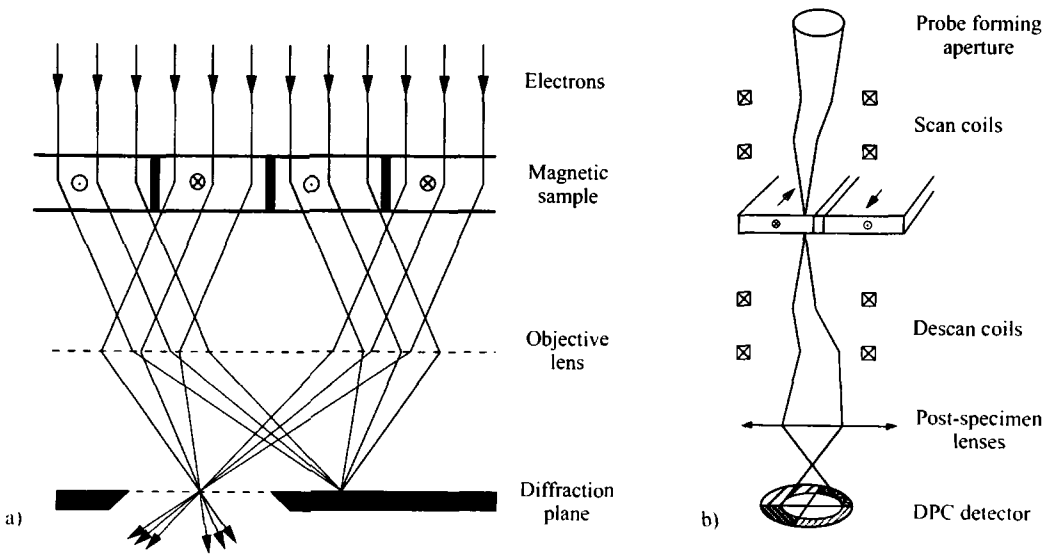


Figure 16. (a) The principles of the Foucault or schlieren method of transmission electron microscopy and (b) differential scanning electron microscopy. In (b) a split detector is employed. The difference signal proves to be a linear function of the magnetic deflection.

still used for detailed investigations of domain walls and inhomogeneities in the domain magnetization. Unfortunately, it is not a high resolution technique and the connection between the obtained images and the magnetic structure is indirect. The same is true for the shadow or schlieren method [38], also known as the Foucault technique. Here, the part of the imaging bundle coming from one class of domains is obstructed in the aperture plane of the microscope, so that mainly electrons from domains of opposite magnetization contribute to the image (Fig. 16a). The result resembles pictures taken with the magnetooptical Kerr effect. A disadvantage of this method is that details of the diaphragm edge determine the imaging of the details of the transition between domains. Advances in microfabrication techniques have led to a revival of this in-focus technique [39]. Particularly attractive is a variant which makes use of modern, highly coherent electron sources. Images taken with this so-called coherent Foucault technique [40, 39] closely resemble images taken with electron holography [41, 42], another technique which combines very clear pictures with high demands on the equipment and on the suitability of the samples.

A truly quantitative and high resolution technique was developed based on the scanning transmission electron microscope [43, 44]. The directional selectivity needed in magnetic microscopy is achieved by a differential detector, as shown in Fig. 16b. The method was called DPC (differential phase contrast) by its inventors based on the mentioned correspondence between electron deflection and phase effects. A field emission electron gun is recommended to achieve the necessary beam quality. It is also useful to discard the central part of the electron bundle and to use annular detectors [45].

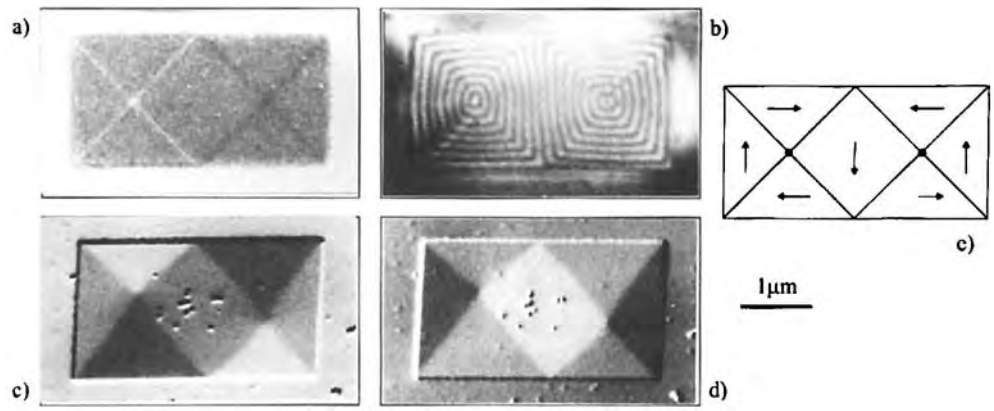


Figure 17. The same type of domain pattern in a thin film element made visible with various techniques: (a) the Fresnel or defocusing technique, (b) the coherent Foucault technique or in-line holography, indicating lines of elementary flux, (c, d) the differential scanning or differential phase contrast method displaying two orthogonal magnetization components of the same pattern and (e) schematic diagram. (Courtesy of S. McVitie, J. Chapman, Glasgow.)

Excellent results can be obtained with this method. The image contrast is linear in the magnetic flux components and quantitative down to at least 100-nm resolution. Its problem is expensive equipment and long image collection times. Usually an overview of a situation is obtained, for example, with the Foucault or Fresnel method, followed by a detailed analysis with the DPC method. Figure 17 shows a comparison of various methods [46].

An interesting variant which promises to offer the advantages of the DPC method without its drawbacks was proposed by Daykin and Petford-Long [47]. It is based on the reciprocity principle between scanning and conventional microscopy, using a series of Foucault images which are digitally combined to become (as claimed) equivalent to the differential phase contrast method. Generally, digital image processing is playing an increasing role in electron microscopy, thus following the example of magnetooptical microscopy.

15.4 Information Storage Media and Mechanical Scanning Methods

15.4.1 Optical Investigation of Magnetic Storage Media

The classical carriers of magnetic recording are particulate media, consisting of small single domain particles embedded in an organic binder. No characteristic magnetic domains can be observed in such materials, and magnetic microscopy can contribute little to their analysis. Things differ with the modern thin film recording media that are always used in magnetooptic recording, which are studied for future vertical recording schemes, and which are increasingly used in conventional longitudinal recording. Anyway, micromagnetic studies on recording media are difficult because magnetic recording relies on nonequilibrium patterns which must be stabilized

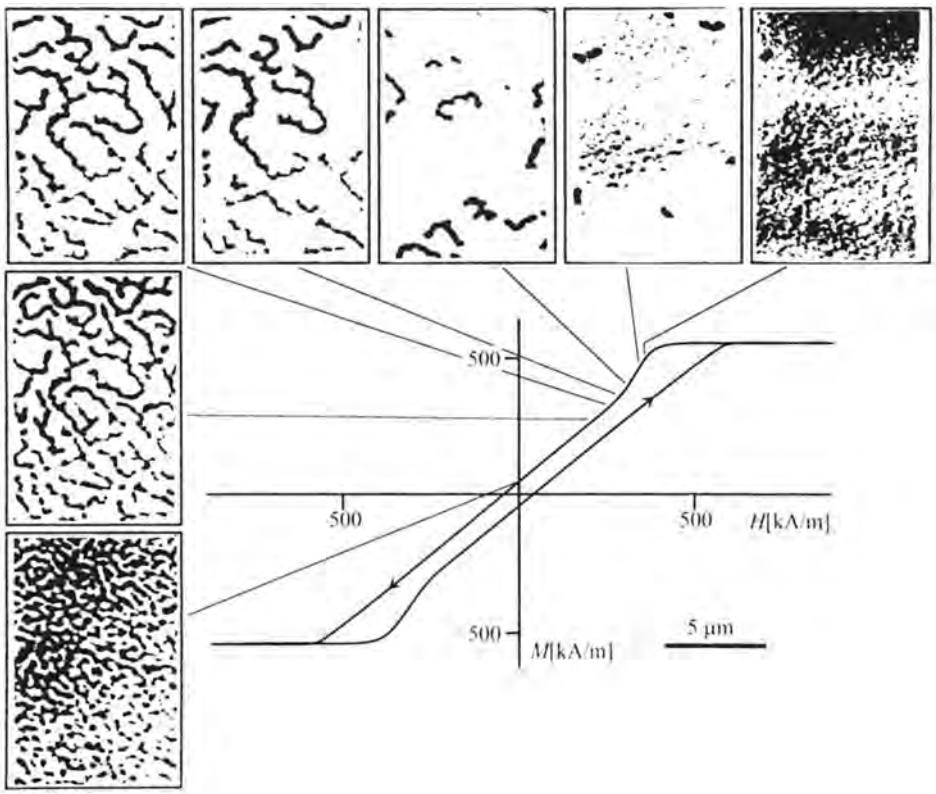


Figure 18. Magnetooptical observation of the magnetization process in a CoCr (17 at. %) perpendicular recording medium (thickness 530 nm). Starting from saturation, opposite domains are first nucleated, then strip-out until they fill the whole sample close to the demagnetized state.

by corresponding media inhomogeneities. Optical observation methods can offer only limited results; other methods are often better suited and tend to largely replace magnetooptics in this area. Some examples of optical observations on recording media may, however, still be reported.

The first observation of magnetic domains in CoCr perpendicular media was achieved by the magnetooptical digital difference technique [48]. The observed patterns approached the limit of optical resolution, but a number of important features could be seen, in particular because the method can be applied at arbitrarily high magnetic fields (Fig. 18). Better resolution could later be achieved by the classical Bitter colloid technique when combined with scanning electron microscopic evaluation [49, 50].

Magnetooptical recording media seem to be natural candidates for optical investigation. This is true to a limited extent. In recording, the numerical aperture of the writing and reading optics is limited to about half the value available in microscopy. More details can therefore be seen in the latter, as demonstrated for example in Weller et al. [51]. In particular, the thermally activated magnetization processes observed under high applied fields in these media can be well studied

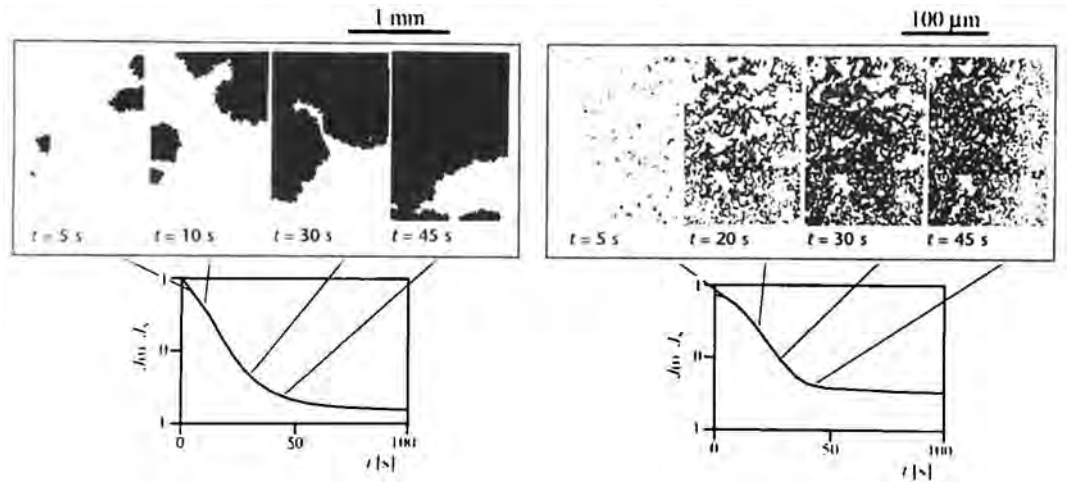


Figure 19. Domain creeping processes in optical recording media consisting of amorphous thin films with perpendicular easy axis. The vertical magnetization component is recorded and displayed by the polar magnetooptical Kerr effect. The character of the reverse domain formation depends on material parameters such as the saturation magnetization, the film thickness, and the coercivity. (a) $Tb_{20.7}Fe_{67.4}Co_{11.9}$, an example of a material with few domain nuclei and a predominance of domain growth processes. (b) $Tb_{26.8}Fe_{56}Co_{17.2}$, a material with many nuclei and dendritic domain growth.

using optical methods, as shown in Fig. 19 [52]. Such studies contribute to a deeper understanding of the material.

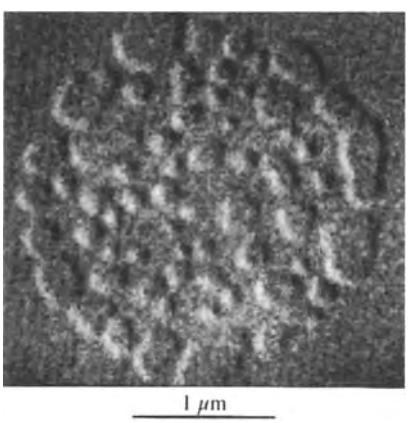


Figure 20. Written spot in a TbFe magnetooptical recording medium, made visible by Fresnel mode transmission electron microscopy. The mark was written in an insufficient magnetic field of only $80\,A\,cm^{-1}$, resulting in an assembly of tiny subdomains instead of the one compact domain needed in recording applications. (Courtesy of C.-J. F. Lin, IBM Research, Almaden, CA.)

In actual recording patterns, details are often hidden below the optical resolution limit, and these details, which contribute to the noise, can only be explored by high resolution methods such as electron microscopy or magnetic force microscopy [53–55]. An example of such observations on written tracks is shown in Fig. 20 [56]. Depending on the writing power, the spots contain an internal substructure which is invisible in optical microscopy.

15.4.2 The Potential of Magnetic Force Microscopy

Impressive high resolution pictures have been produced by the magnetic force microscopy (MFM) of written recording tracks. As this technique is discussed in a separate chapter of this Handbook, we will restrict ourselves here to an evaluation of

the method in comparison with competing techniques.

MFM is sensitive to the magnetic stray field above a sample. In most applications this is a drawback, but not for recording media that in most cases rely on the stray field for readout.

The exact contrast mechanism is disputed; it depends on the details of the probe and of the operation mode. The most confusing aspect of MFM is the interaction of the magnetic probe with the sample. In soft magnetic materials this leads to hard-to-analyze phenomena which are still under active study [57]. This danger is very small for recording materials owing to the intrinsically hard nature of these materials.

Mechanical scanning techniques cannot compete with optical or electron direct imaging techniques with respect to speed. This is not a disadvantage in the investigation of written information so long as the recording pattern and not the recording process is of interest.

The technique can work even through coating layers and on arbitrary substrates. Its resolution is much better than optical resolution, and is comparable with electron optical techniques which need expensive preparations to make them applicable.

Altogether, magnetic force microscopy is developing into the method of choice for the analysis of magnetically recorded information. Figure 21 shows an example of written tracks observed with this technique.

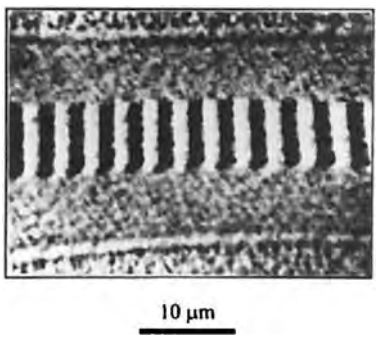


Figure 21. Written track pattern observed in a magnetic force microscope. The periodic pattern is flanked on both sides by a zone in which the previous pattern has been erased prior to writing. (Courtesy of D. Rugar, IBM Research, Almaden, CA.)

15.5 Further Problems and Methods

In the following sections a number of special problems and available methods not covered by the three main fields are reviewed. The selected references should give access to the respective fields rather than being exhaustive.

15.5.1 Domains Under Nontransparent Functional Coatings

In some particularly technical applications, magnetic materials are coated with opaque layers for functional reasons. Electrical steel sheets, for example, must be electrically isolated from each other. In grain-oriented transformer steel, the coating is often used to transmit mechanical stresses. Mineral coatings of a few micrometers thickness are used for this purpose. These coatings preclude the application of the conventional methods of domain observation discussed above. A strong need to get an idea of the micromagnetic processes, nevertheless, led to the development of special techniques.

Most important among these is a special variant of the scanning electron microscope. It relies on the backscattered electrons, and on an oblique incidence of the electrons. Sufficiently high energy electrons penetrate the coating layer and 'see' the magnetic microstructure with reasonable resolution. An introduction to the technique and its applications can be found in the literature [58–61].

As the experimental cost for high-energy scanning electron microscopy is not negligible, many simpler methods have been proposed and applied over the years. Worth mentioning are field scanning methods [62–64] and modified methods based on magnetic particles [65, 66] or even on magnetic bacteria [67, 68].

15.5.2 Magnetic Surfaces and Interfaces

The study of magnetic surfaces and interfaces has attracted increasing interest in recent years. While magnetooptical methods render information from a depth corresponding to the (amplitude) penetration depth of the light (typically 20 nm for metals), there are other methods that are able to probe the very surface on an atomic scale. The already mentioned polarized electron scanning method (Sec. 15.3.2) belong to this category, but there are also other, less developed methods with a strong potential. One is an electron microscope based on spin-polarized low-energy electron diffraction (SPLEEM) [69–71], which is a quite demanding technique but promises much higher image acquisition speeds compared to scanning methods. Another interesting new development is based on nonlinear optics. High intensity laser beams generate second harmonics depending on the symmetry of the scattering medium. The surface or interfaces in layered media are particularly effective in second harmonics, and if the scattered light is analyzed in a similar way as in conventional magnetooptics, magnetic surface and interface structures can be detected [72, 73].

15.5.3 Element-Specific Microscopy

In complex magnetic multilayers it would be interesting to be able to observe magnetic structures in any of the elementary layers. This becomes possible with element-specific X-ray scattering effects. Phenomena analogous to the conventional magnetooptic effects, which are based on the valence electrons, are also present for core electron states and can be detected by X-ray spectroscopy. The effects are associated in this case with narrow core levels which are element specific [74–76].

15.5.4 Methods for Domain Investigation in the Bulk

How to explore the magnetic microstructure in the interior of bulk magnetic materials is a largely unsolved problem. Some quite limited techniques are available, but in most cases the observation of magnetic domains on the surface combined with theoretical arguments for their

continuation into the interior remains the preferred approach.

There are a few direct methods for domain observation in the bulk. X-ray diffraction topography, particularly if based on synchrotron radiation sources, can image the domains due to the slight magnetostrictive lattice distortions usually connected with the magnetic microstructure. Most applications are focused on transformer steel materials, undeformed single crystal sheets of about a tenth of a millimeter in thickness [77–80]. The high cost of this technique together with its limited applicability has not favored its widespread use.

Neutron topography is based on similar principles and is able to penetrate even thicker material. The method has been demonstrated for various ferromagnetic and antiferromagnetic samples [81–84], but the relatively poor resolution and high picture acquisition times cast doubts about its usefulness for ferromagnetic materials (as opposed to antiferromagnets, where it is very valuable due to the lack of other methods).

A unique technique was demonstrated by Libovický [85]. In high-silicon iron alloys an annealing treatment leads to the precipitation of certain oriented particles which are aligned along the local magnetization direction. This structure can later be studied in detail metallographically by dissecting and etching the sample.

Another possible approach is based on neutron depolarization [86, 87]. If the prospect could be realized for this technique to be evaluated in a model-independent, computer-tomographic fashion [88], it would offer an interesting new possibility.

15.5.5 Magnetooptical Microfield Detection

Using magnetooptic indicator films in order to detect magnetic fields above materials that do not display magnetooptic effects themselves has a long tradition in the microscopy of magnetic structures of superconductors. The method relies on induced polar magnetooptic effects generated by perpendicular fields emerging from the sample. Such fields are also present in many magnetic applications connected with hard magnetic materials and magnetic recording. Originally, cerium-doted glass indicators were used for low resolution investigations [89]. Much better resolution could be achieved with evaporated paramagnetic europium chalcogenide films [90–93]. Ferrimagnetic garnet films similar to those used for bubble memories but with an in-plane anisotropy have the advantage of high contrast and sensitivity [94, 95]. An overview of these techniques as applied to superconductors is given in Koblischka and Wijngaarden [96].

15.5.6 Observing Magnetic Structures in Superconductors

Superconductors have magnetic microstructures which are in many respects analogous to ferromagnetic domains [18]. They consist of normal and superconducting domains and can be observed, for example, by the techniques mentioned in the last subsection. Of a special nature are magnetic structures in so-called Type II superconductors which consist of elementary flux lines, the dimensions of which are smaller than the optical resolution limit.

As most technically important superconducting materials are such type II materials, the development of techniques for the observation of flux lines has always attracted attention. The first direct observation of the flux line lattice was achieved by a modified powder technique in which ferromagnetic material is evaporated in a moderate helium vacuum forming a metal particle smoke which decorates the emerging flux lines [97–99].

The renewed interest in superconducting materials stimulated by the discovery of high temperature superconductivity also stimulated the development of new methods which go beyond simply decorating the flux lines. The most direct new method is based on scanning tunneling microscopy. It maps the superconducting electron energy gap and is able to image both flux line lattices and the fine structure of flux lines [100–104]. Another new approach is based on transmission electron microscopy and in particular on electron holography [42, 105–106]. With this method the magnetic flux emerging from flux lines or flux tubes can be displayed directly and very sensitively, and even the dynamics of such magnetic structures can be followed by employing computer-aided image processing.

15.6 References

[1] L. D. Landau, E. Lifshitz, *Phys. Z. Sowjetunion* **1936,** *8*, 153.
[2] F. Bitter, *Phys. Rev.* **1931,** *38*, 1903.
[3] L. v. Hámos, P. A. Thiessen, *Z. Phys.* **1931,** *71*, 442.
[4] F. Bitter, *Phys. Rev.* **1932,** *41*, 507.
[5] L. Wenzel, V. Kamberský, A. Hubert, *Phys. Status Solidi A* **1995,** *151*, 449.
[6] J. Kranz, A. Hubert, *Z. Angew. Phys.* **1963,** *15*, 220.
[7] F. Schmidt, W. Rave, A. Hubert, *IEEE Trans. Magn.* **1985,** *21*, 1596.
[8] W. Rave, R. Schäfer, A. Hubert, *J. Magn. Magn. Mater.* **1987,** *65*, 7.
[9] W. Rave, A. Hubert, *IEEE Trans. Magn.* **1990,** *26*, 2813.
[10] W. Rave, P. Reichel, H. Brendel, M. Leicht et al., *IEEE Trans. Magn.* **1993,** *29*, 2551.
[11] G. Traeger, L. Wenzel, A. Hubert, *Phys. Status Solidi A* **1992,** *131*, 201.
[12] R. Schäfer, *J. Magn. Magn. Mater.* **1995,** *148*, 226.
[13] A. Hubert, M. Rührig, *J. Appl. Phys.* **1991,** *69*, 6072.
[14] S. J. Hefferman, J. N. Chapman, S. McVitie, *J. Magn. Magn. Mater.* **1991,** *95*, 76.
[15] J. McCord. A. Hubert, G. Schröpfer, U. Loreit, *IEEE Trans. Magn.* **1996,** *32 (Intermag '96)*.
[16] R. Szymczak, *Acta Phys. Polon. A* **1973,** *43*, 571.
[17] E. Lifshitz, *J. Phys. USSR* **1944,** *8*, 337.
[18] A. Hubert, *Phys. Status Solidi* **1967,** *24*, 669.
[19] R. Bodenberger, A. Hubert, *Phys. Status Solidi A* **1977,** *44*, K7.
[20] M. Polcarová, *IEEE Trans. Magn.* **1969,** *5*, 536.
[21] J. E. A. Miltat, *Phil. Mag.* **1976,** *33*, 225.
[22] M. Schlenker, J. Linares-Galvez, J. Baruchel, *Phil. Mag. B* **1978,** *37*, 1.
[23] J. Baruchel, S. B. Palmer, M. Schlenker, *J. Phys.* **1981,** *42*, 1279.
[24] R. Schäfer, W. K. Ho, J. Yamasaki, A. Hubert, F. B. Humphrey, *IEEE Trans. Magn.* **1991,** *27*, 3678.
[25] A. Hubert, *Z. Angew. Phys.* **1971,** *32*, 58.
[26] G. S. Krinchik, O. M. Benidze, *Sov. Phys. JETP* **1974,** *40*, 1081.
[27] R. Schäfer, M. Rührig, A. Hubert, *IEEE Trans. Magn.* **1990,** *26*, 1355.
[28] J. Unguris, D. T. Pierce, A. Galejs, R. J. Celotta, *Phys. Rev. Lett.* **1982,** *49*, 72.
[29] K. Koike, K. Hayakawa, *Appl. Phys. Lett.* **1984,** *45*, 585.
[30] J. Unguris, G. Hembree, R. J. Celotta, D. T. Pierce, *J. Magn. Magn. Mater.* **1986,** *54–57*, 1629.
[31] H. P. Oepen, J. Kirschner, *J. Phys. Colloques* **1988,** *49, C8*, 1853.
[32] R. Allenspach, M. Stampanoni, A. Bischof, *Phys. Rev. Lett.* **1990,** *65*, 3344.
[33] M. R. Scheinfein, J. Unguris, D. T. Pierce, R. J. Celotta, *J. Appl. Phys.* **1990,** *67*, 5932.
[34] H. Matsuyama, K. Koike, *J. Elect. Microsc.* **1994,** *43*, 157.
[35] M. R. Scheinfein, J. Unguris, J. L. Blue, K. J. Coakley et al., *Phys. Rev. B* **1991,** *43*, 3395.

[36] M. Speckmann, H. P. Oepen, H. Ibach, *Phys. Rev. Lett.* **1995,** *75*, 2035.
[37] M. E. Hale, H. W. Fuller, H. Rubinstein, *J. Appl. Phys.* **1959,** *30*, 789.
[38] H. Boersch, H. Raith, *Naturwissenschaften* **1959,** *46*, 574.
[39] A. B. Johnston, J. N. Chapman, *J. Microsc. Oxford* **1995,** *179*, 119.
[40] J. N. Chapman, A. B. Johnston, L. J. Heyderman, *J. Appl. Phys.* **1994,** *76*, 5349.
[41] A. Tonomura, *Phys. Rev. B.* **1982,** *25*, 6799.
[42] A. Tonomura, *Surface Sci. Rep.* **1994,** *20.7–8*, 317.
[43] J. N. Chapman, G. R. Morrison, *J. Magn. Magn. Mater.* **1983,** *35*, 254.
[44] J. N. Chapman, *J. Phys. D: Appl. Phys.* **1984,** *17*, 623.
[45] J. N. Chapman, I. R. McFadyen, S. McVitie, *IEEE Trans. Magn.* **1990,** *26*, 1506.
[46] J. N. Chapman, R. P. Ferrier, L. J. Heyderman, S. McVitie et al., *Inst. Phys. Conf. Ser.* **1993,** *138*, 1.
[47] A. C. Daykin, A. K. Petford-Long, *Ultramicroscopy* **1995,** *58*, 365.
[48] F. Schmidt, A. Hubert, *J. Magn. Magn. Mater.* **1986,** *61*, 307.
[49] J. Šimšová, R. Gemperle, J. C. Lodder, J. Kaczér et al., *Thin Solid Films* **1990,** *188*, 43.
[50] J. Šimšová, R. Gemperle, J. C. Lodder, *J. Magn. Magn. Mater.* **1991,** *95*, 85.
[51] D. Weller, R. F. C. Farrow, J. E. Hurst, H. Notarys et al., *Opt. Memory Neural Networks* **1994,** *3*, 353.
[52] S. Winkler, W. Reim, K. Schuster, *Thin Solid Films* **1989,** *175*, 265.
[53] J. C. Suits, R. H. Geiss, C. J. Lin, D. Rugar, A. E. Bell, *J. Appl. Phys.* **1987,** *61*, 3509.
[54] J. C. Suits, D. Rugar, C. J. Lin, *J. Appl. Phys.* **1988,** *64*, 252.
[55] H. W. van Kesteren, A. J. den Boef, W. B. Zeper, J. H. M. Spruit et al., *J. Appl. Phys.* **1991,** *70*, 2413.
[56] D. Rugar, C.-J. Lin, R. Geiss, *IEEE Trans. Magn.* **1987,** *23*, 2263.
[57] W. Rave, L. Belliard, M. Labrune, A. Thiaville, J. Miltat, *IEEE Trans. Magn.* **1994,** *30*, 4473.
[58] B. Fukuda, T. Irie, H. Shimanaka, T. Yamamoto, *IEEE Trans. Magn.* **1977,** *13*, 1499.
[59] W. G. Morris, J. W. Shilling, D. R. Fecich, P. Rao, *IEEE Trans. Magn.* **1978,** *14*, 14.
[60] E. Mivehchi, P. Beckley, D. H. Horrocks, *J. Magn. Magn. Mater.* **1992,** *112*, 58.
[61] T. Nozawa, M. Mizogami, H. Mogi, Y. Matsuo, *IEEE Trans. Magn.* **1996,** *32*, 572.
[62] J. Kaczér, R. Gemperle, *Czech. J. Phys.* **1956,** *6*, 173.
[63] W. Hagedorn, H. H. Mende, *J. Phys. E: Sci. Instrum.* **1976,** *9*, 44.
[64] H. Pfützner, C. Bengtsson, A. Leeb, *IEEE Trans. Magn.* **1985,** *21*, 2620.
[65] H. Pfützner, *IEEE Trans. Magn.* **1981,** *17*, 1245.
[66] G. C. Rauch, R. F. Krause, C. P. Izzo, K. Foster, W. O. Bartlett, *J. Appl. Phys.* **1984,** *55*, 2145.
[67] K. Futschik, H. Pfützner, A. Doblander, P. Schönhuber et al., *Physica Scr.* **1989,** *40*, 518.
[68] G. Harasko, H. Pfützner, K. Futschik, *IEEE Trans. Magn.* **1995,** *31*, 938.
[69] H. Pinkvos, H. Poppa, E. Bauer, J. Hurst, *Ultramicroscopy* **1992,** *47.4*, 339.
[70] K. Grzelakowski, T. Duden, E. Bauer, H. Poppa, S. Chiang, *IEEE Trans. Magn.* **1994,** *30*, 4500.
[71] H. Poppa, E. Bauer, H. Pinkvos, *MRS Bull.* **1995,** *20.10*, 38.
[72] J. Reif, J. C. Zink, C.-M. Schneider, J. Kirschner, *Phys. Rev. Lett.* **1991,** *67*, 2878.
[73] H. A. Wierenga, M. W. J. Prins, D. L. Abraham, T. Rasing, *Phys. Rev. B* **1994,** *50*, 1282.
[74] B. P. Tonner, D. Dunham, J. Zhang, W. L. O'Brien et al., *Nucl. Instrum. Methods Phys. Res. A* **1994,** *347*, 142.
[75] C. M. Schneider, J. Kirschner, *Crit. Rev. Solid State Mater. Sci.* **1995,** *20*, 179.
[76] N. V. Smith, H. A. Padmore, *MRS Bull.* **1995,** *20.10*, 41.
[77] M. Polcarová, A. R. Lang, *Appl. Phys. Lett.* **1962,** *1*, 13.
[78] J. E. A. Miltat, *Phil. Mag.* **1976,** *33*, 225.
[79] B. K. Tanner, M. Safa, D. Midgley, J. Bordas, *J. Magn. Magn. Mater.* **1976,** *1*, 337.
[80] A. Mihara, *Jpn. J. Appl. Phys. Pt. 1* **1992,** *31*, 1793.
[81] M. Schlenker, C. G. Shull, *J. Appl. Phys.* **1973,** *44*, 4181.
[82] M. Schlenker, J. Baruchel, *J. Appl. Phys.* **1978,** *49*, 1996.
[83] J. Baruchel, S. B. Palmer, M. Schlenker, *J. Phys.* **1981,** *42*, 1279.
[84] B. K. Tanner, J. Baruchel, J. S. Abell, *J. Magn. Magn. Mater.* **1992,** *104–107*, 317.
[85] S. Libovický, *Phys. Status Solidi A* **1972,** *12*, 539.
[86] M. T. Rekveldt, *Z. Phys.* **1973,** *259*, 391.
[87] N. Stüsser, M. T. Rekveldt, T. Spruijt, *J. Magn. Magn. Mater.* **1984,** *43*, 73.
[88] M. Hochhold, H. Leeb, G. Badurek, *J. Magn. Magn. Mater.* **1996,** *157/158*, 575.
[89] J. D. Livingston, W. DeSorbo in *Superconductivity* (Ed.: R. D. Parks), Marcel Dekker, NY, **1969**.
[90] H. Kirchner, *Phys. Lett. A* **1968,** *26*, 651.
[91] H. Kirchner, *Phys. Status Solidi A* **1971,** *4*, 531.

[92] A. A. Polyanskii, M. V. Indenbom, V. I. Nikitenko, Y. A. Osipian, V. K. Vlasko-Vlasov, *IEEE Trans. Magn.* **1990,** *26*, 1445.
[93] T. Schuster, M. R. Koblischka, B. Ludescher, N. Moser, H. Kronmüller, *Cryogenics* **1991,** *31*, 811.
[94] L. A. Dorosinkii, M. V. Indenbom, V. I. Nikitenko, Y. A. Ossip'yan et al., *Physica C* **1992,** *203*, 149.
[95] L. H. Bennett, R. D. McMichael, L. H. Swartzendruber, S. Hua et al., *Appl. Phys. Lett.* **1995,** *66*, 888.
[96] M. R. Koblischka, R. J. Wijngaarden, *Supercond. Sci. Technol.* **1995,** *8*, 199.
[97] H. Träuble, U. Essmann, *Phys. Status Solidi* **1966,** *18*, 813.
[98] U. Essmann, *Physica* **1971,** *55*, 83.
[99] R. P. Huebener, *Magnetic flux structures in superconductors*. Springer Verlag, Berlin, Heidelberg, NY **1979**.
[100] H. F. Hess, R. B. Robinson, R. C. Dynes, J. M. Valles, Jr., J. V. Waszczak, *Phys. Rev. Lett.* **1989,** *62*, 214.
[101] H. F. Hess, R. B. Robinson, R. C. Dynes, J. M. Valles, J. V. Wasczczak, *J. Vac. Sci. Technol. A* **1990,** *8*, 450.
[102] H. F. Hess, C. A. Murray, J. V. Waszczak, *Phys. Rev. Lett.* **1992,** *69.14*, 2138.
[103] C. Renner, A. D. Kent, P. Niedermann, Ø. Fischer, F. Lévy, *Phys. Rev. Lett.* **1991,** *67*, 1650.
[104] I. Maggio-Aprile, C. Renner, A. Erb, E. Walker, Ø. Fischer, *Phys. Rev. Lett.* **1995,** *75*, 2754.
[105] T. Matsuda, S. Hasegawa, M. Igarashi, T. Kobayashi et al., *Phys. Rev. Lett.* **1989,** *62*, 2519.
[106] T. Matsuda, A. Fukahara, T. Yoshida, S. Hasegawa et al., *Phys. Rev. Lett.* **1991,** *66*, 457.

Part II

Special Topics

Handbook of Microscopy
S. Amelinckx, D. van Dyck,J. van Landuyt , G. van Tendeloo

1 Small Particles

1.1 Introduction

Small particles play a crucial role in many fields such as catalysis [1], nanophase or nanocrystalline materials [2], clusters [3], and magnetic recording materials [4]. They are applied or occur everywhere in our environment, for example smoke, aerosols, colloidal suspensions, geological erosion, paints, coatings, and materials for grinding and cleaning. Each material and each field of application requires its own type of investigation, which can vary so much that it will be impossible to discuss these within the scope of this contribution, even more so because many different characterization techniques can be applied. Therefore, the emphasis in this contribution is on the specific features and difficulties in the characterization by microscopy techniques and in particular by high-resolution electron microscopy (HREM). Most examples and information given are part of the characterization of catalysts because we are most experienced in this field. The examples are, however, good illustrations for the problems of specimen preparation and image interpretation encountered in the characterization of other small particles. The electron microscope images presented in this contribution were taken with a Philips CM30ST-FEG, except for Figs 11 and 14, which were taken with a JEOL 200CX.

Table 1 lists the type of information needed for a limited number of groups of small-particle materials. A number of examples are given to provide an impression of the relation between the properties, applications, and the characterization methods required. In Section 1.2, relevant characterization methods are briefly given. Section 1.3 give some information on a number of small particle materials. The required information is outlined in Section 1.4. Sections 1.5–1.8 deal with aspects of imaging in transmission electron microscopy (TEM), specimen preparation and specimen handling which are related to small-particle characterization.

1.2 Characterization Methods

1.2.1 Use of Scanning Electron Microscopy, Transmission Electron Microscopy and Scanning Probe Microscopy

The characterization of small particles requires techniques with a high spatial resolution. This requirement implies that

Table 1. Schematic classification of small-particle materials, with wom examples and the most common characterization methods used.

	Catalysts	Dense nanophase materials	Open nanophase materials	Precipitates	Isolated particles	Magnetic materials	Section in which discussed
Examples	$Ni-MoS_2$ (hydrotreating) [5] Pt–Rh (exhaust gas) [11]	Cu (larger hardness) [6] TiO_2 (less brittle) [12]	Ni–Al oxide (catalysis) [7] Al_2O_3 (membranes) [13]	Mg–Si in Al (strength) [8] Y_2O_3 in $YBa_2Cu_3O_7$ (flux pinning) [14]	Er in Al_2O_3 (luminescence) [9] H bubbles in Si (solar cells) [15]	Fe–Cr thin films (recording) [10] Co–Cr particles (recording) [16]	
Characterizations required							
Particle							
Size/density	++[a]	+	+	+	++[b]	+	1.5.1
Shape	+	+	+	++[c]	0	+	1.5.1
Structure	0	+	0	++[c]	++	+	1.5.2
Composition	0	+	0	+	+	++[d]	1.5.3
Surface							
Structure	++[e]	0	++	–	–	–	1.5.4
Composition	++[f,g]	0	++[h]	–	–	–	1.5.5
Grain boundary							
Structure	–	++[i]	++[j]	++[k]	–	–	1.5.6
Composition	–	++[l]		++	–	–	1.5.7
Orientation relation	–	++	++	++[c]	–	–	1.5.6
Ageing	++[m]	++	++	++[k]	++[n]	++	1.5.8

++, very important; +, important; 0, less important; –, not important.

[a] The activity and the selectivity of many catalytic reactions [17] are very sensitive to the size of the particles [18].

[b] Small semiconductor particles, for instance in SiO_2 show photoluminescence with a high efficiency. The luminescence is caused by recombination of exitons confined in the nanocrystals, in which the exitons act as quantum dots, which show a size dependence. For instance, cadmium phosphide, which is black in the bulk material, can be made in all colors of the visible spectrum by varying the particle size between 2 and 10 nm [19].

[c] The effectiveness of the precipitates in the improvement of the properties (e.g., thermal expansion, wear resistance, high elastic modulus and high strength) depend on their structure, their orientation in the matrix and their shape and size.

[d] For instance, the giant magnetoresistance of Fe–Cr depends strongly on the thicknesses of the magnetic and nonmagnetic layers [20].

[e] The structure of small catalytic particles is very important. The catalysis requires an adsorption on the surface, the rate of which depends strongly on the presence of surface steps and surface facets.

[f] Promoters are elements added to a catalytic system but which are not catalytically active themselves [21]. They are added to improve the activity, selectivity and/or stability. For instance, the addition of vanadium to $Rh-SiO_2$ results in an increase of the rate of methane formation from CO and H_2O [22]. To understand the promoter effect the location of the promoter atoms on the surface must be known.

[g] The activity and selectivity of catalysts consisting of metal particles on supports which are relatively easy to reduce (e.g., the selectivity of rhodium or platinum on TiO_2 [23]) can change strongly upon reduction at temperatures in the range of 200–500°C. This phenomenon, called strong metal–support interaction (SMSI), is due to a very thin layer (0.2–0.6 nm) of partly reduced substrate (e.g., TiO_{2-x}) covering the metal particles.

[h] In the case of Raney nickel catalysts [24], which are composed of nickel, aluminum and oxygen, the addition of a small amount of, for instance, chromium or iron does result in improved performance for some reactions [25]. These added metals could be dispersed on or close to the surface, thus resulting in a significant change in activity or selectivity.

[i] The hardness of copper consisting of 10 nm particles is five times greater than that of copper consisting of 100 mm grains [26]. This increase in hardness is related to the atomic structure at the grain boundaries, which are abundantly present in the material with 10 nm particles.

[j] Ceramic membranes are used for various applications depending on the pore sizes [27]. Membranes with pores in the size range 1–2 nm are used for instance for gas purification. One of the earliest uses of such membranes was the enrichment of ^{235}U [28]. A major requirement of these membranes is a high thermal stability, which is for a major part determined by the grain boundaries.

[k] Precipitates are often introduced into steels and intermetallic alloys [29] by choosing proper compositions and annealing conditions to improve mechanical properties such as creep resistance and flow strength at high temperatures. The improvement of mechanical properties is due to the pinning of dislocations at precipitates, for instance metal carbides in steel [30], or ordered Ni_3Al in nickel–base alloys [31]. In the latter case, the movement of dislocations in the disordered g phase (f.c.c.) is impeded by the precipitates, since these dislocations are not perfect in the ordered phase, that is, are bounded by high-energy antiphase boundaries or superlattice intrinsic stacking faults. However, if too many dislocations are piled up at the precipitate dislocation, slip can occur, resulting in a poorer performance.

[l] Nanocrystalline alloys consist of a uniform mixture of crystallites with different chemical compositions. However, at the grain boundaries a mixing of the two chemical compositions occurs, resulting in new properties.

[m] Ageing in automobile exhaust catalysts (existing of Pt/Rh with CeO_2 as the promoter and Al_2O_3 as the support) occurs, because of sintering of the metal particles or the CeO_2. In an ageing study, one follows the deactivation of the catalyst in relation to microstructural and chemical changes.

[n] Hydrogen is added to amorphous silicon in photocells to passify defects. At present, up to 10 at.% of hydrogen is incorporated. Most of the hydrogen is dissolved in the amorphous silicon, but hydrogen bubbles can be formed, in particular during extensive testing.

electron microscopy and scanning probe microscopy (SPM) are invaluable tools in the research of small particles. Still other characterization techniques are equally important because they allow bulk characterizations yielding information which should be in agreement with the microscopic information. The techniques most often used for information averaged over the surface are X-ray photoemission spectroscopy (XPS), electron energy loss spectroscopy (EELS), and secondary ion mass spectroscopy (SIMS), whereas bulk information is obtained by Mössbauer spectroscopy, infrared spectroscopy, extended X-ray absorption fine structure (EXAFS), and low-angle X-ray diffraction. These techniques can provide information (e.g., surface composition, relaxation at grain boundaries, and electronic states) which is essential for the understanding of the performance of a given material, and which is difficult or impossible to obtain by the microscopy techniques given above. Thus, microscopy techniques are almost always used in combination with a number of other techniques.

Since the limitations of scanning electron microscopy (SEM) and TEM on one side (a good vacuum is required, allowing no real atmosphere conditions) and STM on the other side (a rather flat specimen is required, making the study of real catalytic materials impossible) are complementary, a combination of both methods can be very fruitful.

1.2.1.1 Scanning Electron Microscopy

SEM is mainly used for (i) the identification and distribution of phases and (ii) size and morphology determination. The applicability of SEM for these determinations is strongly related to the ratio of the particle size to the resolution of the electron microscope. Hirsch et al. [32] report that one can optimize the visibility of particles of a certain size by tuning the acceleration voltage. In particular, for the determination of orientation relations of neighboring grains in dense nanocrystalline materials the use of scanning electron microscope diffraction [33] can be very useful.

Charging is observed for quite a number of small-particle materials, such as catalyst particles on highly porous ceramics. The conventional method of preventing charging – coverage with a layer of metal (usually palladium or gold applied by sputtering, with a thickness ranging between 3 and 30 nm) or carbon – can only be applied if the details one wants to study are not significantly obscured by the coverage. The most straightforward approach to prevent charging without coverage of the specimen is that of operating the scanning electron microscope at an accelerating voltage at which the secondary electron emission coefficient is unity, thereby preserving charge neutrality in the specimen. Most conventional scanning electron microscopes are capable of operation in this regime (1–2 kV); however, as most instruments are designed for optimum performance at around 30 kV the resolving power at 1 kV is considerably less than what is theoretically attainable. Thus, one does not generally reach a resolution which is of interest for the study of small particles with conventional scanning electron microscopes. A different approach to the specimen charging problem is that of surrounding the specimen with a low-pressure atmosphere, such as water

vapor, that is capable of conducting away the induced charge. A third route entails discharging the specimen surface by tracking the electron beam closely with a carefully controlled beam of positive ions. Such a dual-beam instrument can have several additional features that utilize the focused ion beam such as ion milling, etching and deposition.

An advantage of SEM over TEM is that one can sample the surface by tuning the acceleration voltage. This allows the determination of the fraction of catalytic material on the surface. To show the importance of this, consider the distribution of metal particles in a zeolite crystal. A zeolite is an open-framework structure with pores between 0.4 and 1 nm in diameter. The zeolite framework is used to allow only specific catalytic reactions, which are governed by the size of the channels. This shape-selective catalysis requires that as few catalytic particles as possible are located on the outside of the zeolite crystals, which can be checked better with SEM than with TEM. STM is not possible because of the roughness of the zeolite.

1.2.1.2 Transmission Electron Microscopy

The need for TEM investigations can be divided into six areas: (i) identification of phases, (ii) distribution of a certain phase, (iii) size and shape determination, (iv) determination of the atomic arrangement, (v) determination of the local element composition, and (vi) the determination of the local electronic structure.

Phase Identification

Phase identification can be done based on the unit cell, the atomic structure or the composition, using electron diffraction, HREM and element analysis, respectively.

Phase Distribution

The distribution of a certain phase can best be done by dark field imaging (if the phase is crystalline) or by element mapping. The latter one can be realized in the scanning TEM (STEM) mode or by using energy-filtered imaging.

In dark field imaging, only the information from a selected part of the diffraction pattern is used for imaging. This allows selection of an area in diffraction space which contains only information from the small particles of interest (e.g., niobium carbonitride precipitates in ferrite [34]). Frequently the small particles have a random orientation (e.g., platinum particles on alumina [35]), resulting in rings in the diffraction pattern. By selecting only a segment of such a ring, one images only a small fraction of the particles. This problem can be overcome by using an objective aperture which consist of a thin circular slit at the right Bragg angle [36]. This method is, however, very difficult in practice because only one diffraction angle can be applied with a given aperture. A much easier method is to use a tilted beam which is modified electronically such that it describes a hollow cone [37]. The latter method allows selection of any Bragg angle for imaging, and the width of the Bragg angle can be selected with the size of the objective aperture. One should take into account that the smaller the size of the objective aperture the larger the uncertainty in the determination of the particle sizes will be.

Determination of Size and Shape

The ease of determination of the size and

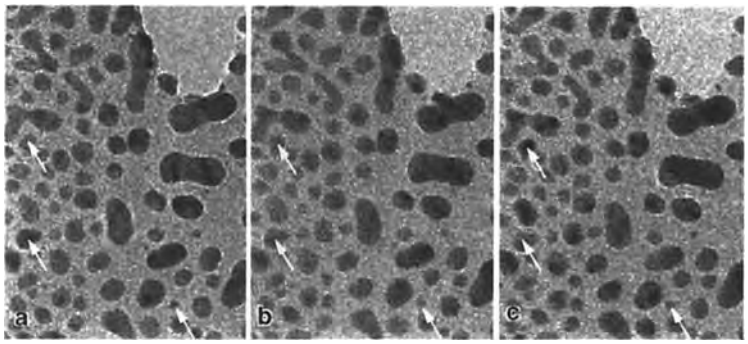


Figure 1. Gray value images of gold particles for three different orientations (0, 5, and 10° tilt), showing the variation of the relative gray values of the particles. The speciment was made by sputtering gold on a thin carbon foil. Some particles showing the change in contrast are indicated by arrows. In this example the support is very thin and amorphous with clearly visible particles, but if the specimen was thicker or (partly) crystalline the part of the particles having a low contrast would be difficult to detect.

the shape depends strongly on the visibility of the particles. If the particle is crystalline, its visibility is strongly influenced by its orientation. If the particle is in an orientation in which it diffracts strongly it is much more visible than in an orientation in which diffraction is only weak. Thus, it can be advantageous to take images at specimen orientations which are rotated with respect to each other (Fig. 1).

Determination of the Atomic Arrangement

The atomic arrangement with a resolution of about 0.17 nm can be obtained through straightforward HREM experiments. The resolution can be further improved by image reconstruction methods up to the information limit of the electron microscope, which is about 0.12 nm for a microscope with a field emission gun.

Determination of the Local Elemental Composition

The local elemental composition can be determined with the standard techniques of EDX and EELs. The only requirement is that for a given small spot size the signal-to-noise ratio is sufficient. Since the smaller the particle to be analyzed the larger the effect of specimen drift, for particles smaller than 5 nm one needs a high electron beam intensity, which can be readily obtained with microscopes equipped with a field emission gun. If one uses such a high electron dose the specimen itself can change (e.g., the irradiated area can be amorphized or removed (hole drilling)). The composition of only the surface and subsurface (depth 1–2 nm) can be determined using Auger electrons. Apart from commercially available scanning Auger microscopes (with a resolution of about 100 nm) this is also possible in specially designed transmission [38] and scanning [39] electron microscopes. This technique is especially interesting for catalysis research, since a spot size less than 5 nm is now obtainable.

Determination of the Local Electronic Structure

The electronic structure of an area of a particle can in principle be determined

from EELS data, since the loss of energy of the inelastically scattered electrons is determined by the electron–specimen interactions, which are not only atom-dependent (coarse structure of EELS spectrum) but depend also on the environment of the atom (fine structure of the EELS spectrum). EELS is presently possible with a resolution of approximately 0.5 eV due to the energy spread of the primary electrons. However, it is to be expected that with a number of adaptations of the electron microscope the resolution can be improved to about 30 meV. This can provide important additional information such as the site and surroundings of a specific atom, the bonding of an atom and/or the electronic structure.

1.2.1.3 Scanning Transmission Electron Microscopy

STEM is used for the same kinds of investigations as outlined above for TEM. With respect to TEM, the strength of STEM is that a variety of signals are available with a spatial resolution of down to about 0.2 nm [40], for example energy loss, characteristic X ray, secondary electron emission, Auger electron emission, annular dark field, and high-angle annular dark field imaging [41]. The versatility of a scanning transmission electron miscroscope, especially in dark field imaging and element mapping, makes it well suited for studies of particles and precipitations. In particular, for the imaging of very small particles STEM can be superior to TEM. Howie showed that by using a high collection angle it is possible to obtain a higher and more reliable contrast from single atoms and randomly oriented clusters than in traditional dark field techniques [42].

When using the normal annular dark field detector, an image based on Bragg scattering can be obtained similar to the TEM hollow cone condition, only the collection efficiency is better in the STEM case. This is normally used when one deals with a particle or an ensemble of particles which can then be 'highlighted'. Other ways of highlighting are by elemental mapping using EELS, EDX or a combination of several of the collected signals.

Alternatively, high-angle incoherent scattering can be exploited to form an image which depends only on the scattering of the projected potential (Z contrast), with a $Z^{3/2}$ dependence. The image formed with this set-up is intuitively interpretable, but the signal-to-noise ratio is inferior to what can be obtained by conventional HREM. High-angle scattering has successfully been used in catalysis to image small metal clusters [43].

Because the signals from STEM detectors are measured electronically, they can be manipulated mathematically. In this way the imaging capability of fine particles can be enhanced by applying ratio techniques [44], thus reducing the contribution of a substrate.

1.2.1.4 Scanning Probe Microscopy

The usefulness of SPM in the characterization of small particles is restricted because of the geometry of the scanning tip and because many technologically important small particles are isolators or are located on isolating supports. The latter restriction applies to scanning tunneling microscopy (STM). Concerning

the first restriction, one requires a scanning tip which is much sharper than the particle itself, because the SPM image is a convolution of the specimen shape with the shape function of the SPM tip. Apart from the small size of the particles, an additional problem can arise from the supporting material. For industrial catalysts this support is highly irregular, making SPM or atomic force microscopy (AFM) impossible. Therefore, only very flat substrates (e.g., SiO_2 on a silicon wafer) are used as model supports in SPM and AFM studies. The studies with these substrates are limited to the determination of the size distributions of the catalytic particles. Ma et al. [45] obtained SPM images of 1–2 nm platinum and nickel clusters on flat graphite surfaces. They observed that some clusters disappear between two scans with a time interval of only 5 s, which they explain by diffusion of those clusters which have only little interaction with the substrate. However, it cannot be excluded that given the small particle–substrate interaction the small metal particles are displaced by the tip.

The sharpness of the tip is crucial in the study of small particles. This has been illustrated by Sheiko et al. [46]. They used a faceted surface of $SrTiO_3$ containing (1 0 1) and (1 0 3) facets with atomically sharp edges. The angles of these facets with the overall surface (3 0 5) was +14 and −13°, respectively, indicating that the change in slope was fairly small. These sharp facets could only be imaged with a small fraction (<10%) of the commercial high-quality Si_3N_4 tips. They proposed that such a $SrTiO_3$ surface can be used to test the quality of a tip. In the case of small metal particles the changes in slope will be much more drastic than in the case of the $SrTiO_3$, setting even stricter limitations to the sharpness of the tip.

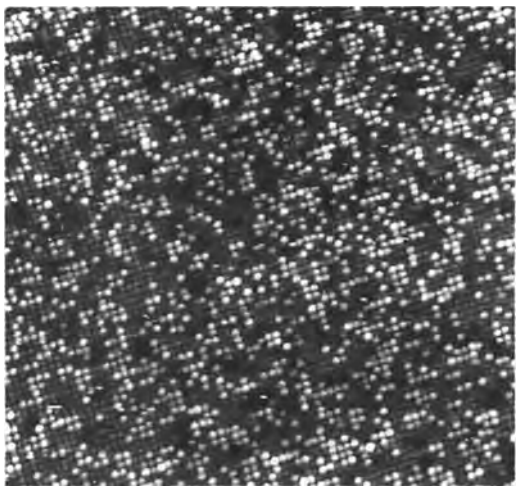


Figure 2. STM image of a flat (1 0 0) Pt–Rh surface. The platinum atoms (white dots) can be distinguished from the rhodium atoms due to a difference in height of 0.02 nm.

STM and AFM are used in the study of catalysts, but in these cases large (single-crystal) metal surfaces are used. These kinds of substrates allow the imaging of individual atoms at the surface or absorbates on the surface, which also allow spectroscopy on single adsorbed molecules [47]. An example of the possibilities of STM on flat metal surfaces is shown in Fig. 2. In this STM image one can distinguish platinum and rhodium atoms on a (1 0 0) surface of a Pt–Rh alloy by a difference in height of 0.02 nm [48]. Another important possibility of STM is the monitoring of surface changes upon a given reaction (e.g., an oxidation) as long as this change is not too fast. Since STM can be done under atmospheric pressure, or even higher pressures, this option allows the study of model materials (e.g., flat surfaces of single crystals) under conditions which are close to the conditions

Table 2. Some characterization methods applied in the analysis of small particles.

Method	Application
Surface-sensitive techniques	
X-ray photoemission spectroscopy [50, 51]	Composition, absorption, atom coordination
Auger spectroscopy [51]	Composition, depth profiling
Low electron energy loss spectroscopy [51]	Composition, atom coordination
Secondary ion mass spectroscopy [51, 52]	Composition
Temperature-programmed techniques [54]	Absorption, decomposition
Bulk techniques	
Mössbauer spectroscopy [55, 56]	Atom coordination
Infrared spectroscopy [57]	Absorbed molecules
Extended X-ray adsorption fine structure [58]	Atom coordination, size
X-ray diffraction, neutron diffraction [59]	Average size, structure
Low-angle diffraction [60]	Size distribution
Whole diffraction pattern analysis [61]	Size distribution, average grain boundary relaxations
Nuclear magnetic resonance [62]	Atom coordination/symmetry

used to perform the catalysis (in which case small particles are used). The third important research subject is the imaging and spectroscopy (scanning tunneling spectroscopy (STS)) of single adsorbed molecules on flat surfaces [49]. In the case of STS the variation in the density of states of the adsorbed atom are probed by varying the size of the energy window, which defines the energy range of the contributing states. The density of occupied and unoccupied states is thus reflected in the tunneling current.

1.2.2 Other Characterization Methods

Of course, many other characterization methods are applied for small particle analysis. In Table 2 a number of methods are given with references to books or review papers. Only the methods most frequently used and their main applications in the characterization of small particles are listed.

1.3 Small-Particle Materials

1.3.1 Catalysts

Catalysts are used in a wide range of processes. In this chapter only the heterogeneous catalysts will be considered. These are for instance used in the petrochemical industry. In catalytic reforming, molecules with a low octane number are converted into molecules with a high octane number. This process is used for more than 10^9 kg per day. One uses a silica–alumina matrix containing zeolite crystals as the catalyst. Hydrotreating in the oil refinery process involves conversions due to reaction with hydrogen, and it is in particular used for the removal of hetero-atoms (e.g. sulfur, nitrogen, and oxygen). One method uses supported mixed phases of, for example, CoS and MoS_2 or NiS and MoS_2 as catalysts.

The heterogeneous catalysts consist usually of two or more components: the support and one or more active phases. The active phase usually constitutes

between 0.1 and 20% of the total catalyst, and is normally in the form of small crystallites (1–50 nm). The support is usually considered to be catalytically inactive by itself. However, in partnership with the active phase it can contribute in an important way to the total reaction. The substrate material should have a high surface area. Typical high surface area substrates are silica and γ-alumina, having surface areas in the range of 100–800 $m^2 g^{-1}$.

Catalyst preparation is still, at least for a significant part a sort of alchemy, in which by trial and error one develops a catalyst which works, without knowing in detail why it works (better than another catalytic material). For instance, the performance of a certain nickel-on-a-silica catalyst can be different from a similar one which has been prepared from a different batch of silica but which should have the same properties according to the manufacturer of the silica. In this section only the most important preparation routes will be outlined.

One preparation method in which the support and active phase are reacted together is based on selective removal of, for instance, water and oxygen. An example of this is an old but still important iron-based catalyst for ammonia synthesis. This is made by melting Fe_3O_4, alumina, and potassium oxide and subsequently removing most of the oxygen by reduction. This yields highly porous and mechanically strong particles.

More often one uses methods of depositing the active phase on a separately produced support. In this case one can use precipitation, impregnation, adsorption and ion exchange. When more than one active component has to be deposited, one can use co-precipitation, which, however, is difficult to control and to reproduce. In all cases the main goal is to distribute as randomly as possible the active atoms or molecules over the substrate. Furthermore, during the process steps towards activation of the catalytic phase(s), a good dispersion should either be maintained or obtained. To reduce the costs of a catalyst one generally aims at such a dispersion that an as large as possible fraction of the atoms of the active phase is at the surface without losing its activity or stability. An example of impregnation is an alumina-supported platinum catalyst. which is made by impregnating a preshaped support with an aqueous solution of H_2PtCl_6; the acid adsorbs strongly on the alumina support. Next the support is dried, calcined and reduced, to remove the water, chlorine and oxygen, respectively.

Another approach to the deposition of small particles is to produce the small particles first with a fixed number of atoms per particle, as has been described by Schmid and co-workers [63]. In this synthesis route one obtains metal particles with a skin formed by large organic groups. The metal core is built from the central atom by surrounding it by a number of close packed layers, yielding distinct numbers of atoms thus forming a series 1, 13, 55, 147, 309, 561, ... The advantage of this route is that only particles with a fixed number of atoms occur, allowing direct correlation between particle size and catalytic performance. An example of palladium clusters is shown in Fig. 17.

Apart from the actual catalytic particles, the morphology of the support can play a large role. For instance, for the catalytic refinement of crude oil the catalyst comprises a well-balanced

conglomerate of particles, such that they have a 'chain' of pores which narrow down from pore sizes of several hundreds of nanometers at the surface to pores of about 1 nm. The conglomerate consists of zeolite Y crystals, silica, γ-Al_2O_3, and platinum metal particles. Zeolite Y contains pores of about 1 nm, and γ-Al_2O_3 and silica are used not only as the binder but also to provide the larger pores. The acid sites in the zeolite and on the γ-Al_2O_3 catalyse cracking (C–C bond breaking) and isomerization (forming different C–C sequences), whereas the metal particles catalyse mainly hydrogenation (adding hydrogen atoms to the carbon chains). Since bond breaking and hydrogenation should preferably occur at approximately the same location it is essential that the platinum particles are finely dispersed. An additional reason for a fine dispersion is that one can reduce the cost of the platinum. If the platinum particles become too small (<1 nm) they are more likely to migrate and cluster into larger ones.

A review on TEM of zeolites is given by Terasaki [64]. Zeolites are open-framework structures, of which the framework consists of $(Si,Al)O_2$ with silicon and aluminum in tetrahedral coordination forming a one-, two-, or three-dimensional array of channels, with diameters from 0.4 to 1 nm. The pore system allows access by and transport of those molecules which fit into the pores. As such, they act as molecular sieves with an invaluable shape selectivity for reactants as well as products. Some zeolites do occur in nature, but most of them are synthetic. The synthesis of zeolites is done under hydrothermal conditions. Many different structures have been synthesized by the use of specific templates, organic molecules which fit in the pores of one zeolite but not in others thus governing the synthesis of particular zeolites. Obviously a major effort has been put into the synthesis of zeolites with a pore diameter significantly larger than 1 nm. Only some limited results have been obtained, such as 1.2–1.3 nm pores in a zeolite called VPI-5 [65]. Recently the limited pore size barrier was broken by the discovery of a new family of ordered mesoporous molecular sieves by researchers at Mobil [66]. These new materials, called MCM, are synthesized under hydrothermal conditions from (alumino-)silicate gels containing organic surfactant molecules such as hexadecyltrimethylammonium chloride (C_{16}TMACl). By the choice of the length of the carbon chain one is able to choose the diameter of the pores, which can range from 1.5 to 10 nm. These materials open the way to the use of shape-selective catalysis to convert molecules which are too large to fit into the zeolite lattices. MCM is formed by the solidification of $(Si,Al)O_2$ around the sphere-like or pillar-like micelles composed of the surfactant molecules, thus forming tubes or plates. MCM can occur in various morphologies [68, 86]: a rather chaotic one, in which tubes are strongly curved and have several interconnections; (ii) a hexagonal stacking of the tubes, of which an example is shown in Fig. 3; (iii) a cubic arrangement of interconnecting tubes; and (iv) a plate-like morphology. The last one is not so interesting because the structure collapses when the template (the surfactant molecules) is removed. The tube walls do not show a periodicity along the direction of the tube and it should be considered as an amorphous component. Apart from the pore size, the thickness of the pore walls can also be manipulated.

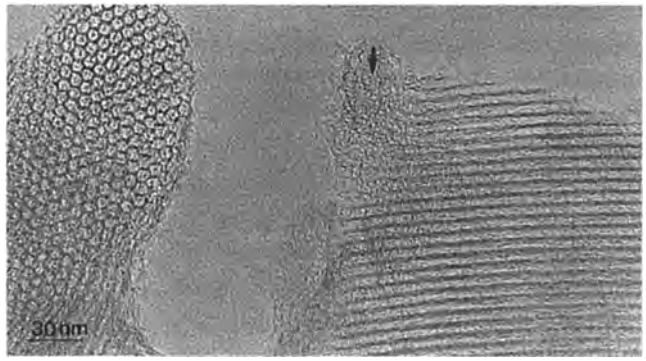


Figure 3. A TEM image of two MCM crystals on an amorphous carbon grid. The MCM has a hexagonal structure with an *a* axis of 7 nm. The crystal on the left is viewed along the hexagonal axis, and shows a sixfold symmetry for the tube arrangement. Note that the bottom part if this crystal is misoriented. This local change in orientation is observed in many MCM crystals. The crystal on the right is viewed perpendicular to the channels. Note that one cannot observe any lattice fringes except those of the channels themselves, indicating that the wall of the tubes is amorphous.

Since heterogeneous catalysis occurs at the surface of the catalytic particles, a description on the atomic level is needed. This requires a very local characterization, in particular because most catalysts are very complex systems. Quite often the catalysts are prepared with the addition of other elements about which it is known only that their presence results in a better activity or selectivity or in an improved stability of the catalyst. Such additives are called promoters. In all cases the ultimate goal is an understanding of why a certain catalyst shows such a good performance, to allow the search for further improvements in a systematic way.

The complexity of a given catalyst makes it necessary in most instances to do HREM investigations on a series of catalysts. For instance, in a recent study [68] of the activity of the hydrodesulfurization of the sulfur-containing molecules thiophene and dibenzothiophene using W–Ni–S on γ-Al_2O_3 it was found that the activities for thiophene and dibenzothiophene were respectively not dependent and quite strongly dependent on the temperature of the preparation of the catalysts (calcination and sulfidation). The most frequently reported HREM feature of these catalysts is the presence of dark contrast bands in the HREM images, which are due to slabs of WS_2 possibly with some incorporation of nickel [69]. In the literature these slabs are reported to be the active component for the hydrodesulfurization. The catalysts with the higher activity for desulfurization of dibenzothiophene were shown, however, to contain very few or no slabs but very small clusters (about 0.5 nm) which were barely visible. These clusters probably contain nickel, tungsten, and sulfur. Catalysts prepared with higher sulfidation temperatures showed no small clusters but only slabs to be present. However, these catalysts still showed a significant activity for desulfurization of dibenzothiophene (about one-third that of the catalyst containing the clusters), indicating that the slabs are catalytically active but less than the small clusters. HREM images of both

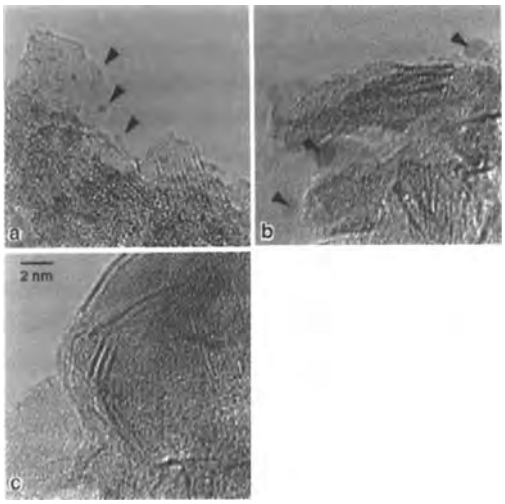


Figure 4. Two HREM images of W–Ni–S on g-Al_2O_3. (a) A specimen obtained by calcination at 650°C and sulfidation at 400°C. Only small black spots can be seen (some of them are indicated with an arrow). (b) A specimen calcined at 400°C and sulfidated at 650°C. Black lines corresponding to WS_2 slabs can be observed throughout the image. For some reactions the small spots in (a), which are barely visible by HREM, show the highest catalytic activity.

catalysts are shown in Fig. 4. The clusters are difficult to distinguish, and it should be noted that due to the high scattering potential of tungsten the imaging conditions are favorable for this catalyst. This example shows that one should take into account the possibility that the most active phase is very difficult to detect by HREM. Local element analysis can be very helpful in this case as can be a careful HREM characterization of a series of catalysts showing different activities.

Another area in which microscopy techniques are used is the study of deactivation (ageing) of a catalyst during use. For instance, during the catalytic cracking of oil the catalyst tends to lose 90% of its activity within 1 s, due to the deposition of a thin carbon layer (coke). An elegant solution has been found to this problem by performing the catalytic cracking in one reactor and the regeneration of the catalyst in another one placed in series in which the deposited coke is burned off. Other deactivation processes involve the sintering of particles, poisoning by other elements (e.g., the deactivation by lead of catalysts for cleaning exhaust gases in cars), and changes in the porosity of the substrate. Two approaches can be used to study the deactivation; in both cases the catalyst is investigated before and after ageing. In the first approach one investigates different specimens; by undertaking a statistical analysis one can reveal the differences. In the second approach one investigates the same areas before and after ageing; this requires more careful specimen handling (see Sec. 1.8.5 of this Chapter) but has the advantage that one can easily determine the changes.

Most catalytic reactions occur at elevated temperatures and pressures. Thus the conditions during the catalytic reaction are much different from the conditions in the electron microscope. This points to two disadvantages of electron microscopy, even under ideal conditions. In the first place one has to take into account that the catalytically active phase can be changed by the vacuum of the microscope. In the second place, reactions cannot be followed in situ (see also Sec. 1.8.4 of this Chapter). For ageing studies, STM can be a very useful tool because it does not require a vacuum.

1.3.2 Nanophase Materials

Nanophase materials consist of synthetic materials with an average grain size below

100 nm [70]. The various grains can be of different phases or structures. The basic idea behind nanophase materials is the generation of materials with new atomic structural arrangements by introducing a very high density of defects such as grain boundaries, interphase boundaries, and dislocations resulting in new macroscopic properties. The density of metallic nanocrystalline materials varies between 75 and more than 90% of the crystalline density. Nanocrystalline materials have also been called ultrafine grain materials, nanophase materials, or nanometer-sized crystalline materials. The interest in these materials is rapidly growing because it has been shown that some of their properties are superior to 'conventional' materials, which are composed of much larger grain sizes [71]. Nanophase materials span a wide range of compounds, for example metals, oxides, nitrides, and semiconductors. The use of these nanophase materials is also very broad. In the case of metals, they show—when compacted to almost fully dense materials—a greater hardness (e.g., the hardness of copper consisting of 10 nm particles is five times greater than that of copper consisting of 100 μm grains [72]. On the other hand, the uncompacted metal can have a very high surface/bulk ratio, making it suitable for catalysis (e.g., Raney nickel [73]). Ceramics (e.g., TiO_2 and ZnO [74]) consisting of very small particles are less brittle than materials with grains of conventional sizes. Small semiconducting particles (e.g., silicon in amorphous SiO_2) can show interesting luminescence properties [75] and can exhibit quantum phenomena. Multilayers, consisting of layers of small particles, can be used as X-ray mirrors (e.g., Pd–C [76]) or can be used to study amorphization phenomena (e.g., Ni–Ti interfaces [77]).

The manufacturing of nanophase materials has been reviewed by Gleiter [78], and can be classified into three groups.

The first group of fabrication techniques is cluster assembly. This again covers a broad field with chemical precipitations from solutions [79] and sputtering (for instance into an oil to separate the particles or onto a substrate to form multilayers). In particular, gas condensation is increasingly used, especially in the preparation of ceramic materials.

The second group is mechanical attrition, which is usually done by ball milling [80]. Using this technique, ultrafine particles can be obtained in a cheap way, but some contamination from the ball milling device is likely to occur.

The third group of methods is implantation by high-energy ions, for example the implantation of erbium into amorphous SiO_2 [81]. By an additional annealing process the deposited ions can cluster, resulting in nanoparticles, the size of which depends on the dose and the annealing conditions. This technique is very expensive, since it requires an accelerator.

Nanocrystalline alloys may be generated by simultaneously evaporating or sputtering several components, which can result under favorable conditions in a uniform mixture of crystallites with different chemical compositions. A major advantage of such materials is that chemically different atoms can be mixed at the grain boundary interfaces [82].

The diversity of compounds and applications is also the actual reason for using electron microscopy techniques.

Evidently, size and shape can be easily determined by TEM, whereas small angle X-ray and neutron diffraction allow determination of the size distribution. In dense nanophase materials the grain boundary structure is the most important part, since this affects the improvements in the mechanical properties such as hardness and brittleness. Determination of the grain boundary structure is, however, an almost impossible task. It requires that the grain boundary is (almost) parallel to the electron beam, without steps in that direction, and is continuous throughout the specimen (Fig. 5). The grains on both sides of the boundary should also be in such an orientation that the structures of both grains are imaged. Finally, the specimen should not be covered with an amorphous layer. An additional problem is the delocalization of information, which is discussed in Sec. 1.5.1 of this Chapter.

Since the boundary core structure depends on the interatomic bonding forces and the boundary crystallography and since many different grain orientations and grain boundary orientations occur, a large variety of grain boundary core structures will occur in a given specimen. Therefore, an HREM study can only provide fragmented information, describing at best only a part of the grain boundary structure. A study of the atomistic structure of grain boundaries in metal nanocrystalline materials by HREM suggested that these grain boundaries are similar to the grain boundaries in conventional polycrystalline materials [83]. In this favorable condition, specimens with larger particles can be used as model specimens for the

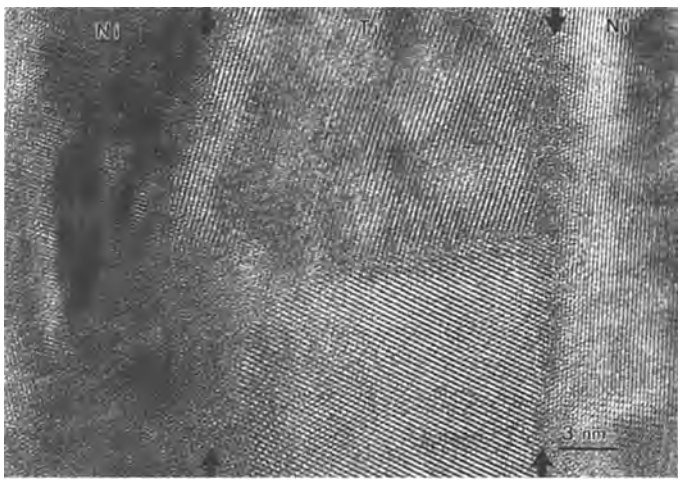


Figure 5. HREM image of a part of a Ni–Ti multilayer, composed of nanocrystalline nickel and titanium. The multilayer is grown from right to left. One can observe a number of grain boundaires, between grains in neighboring nickel and titanium layers and between two grains in the titanium layer. In this example the grain boundaries are almost parallel to the electron beam, but due to misalignment of the crystals the atomic structure of the grain boundary cannot be determined. The chance that both grains are in a low index orientation with the grain boundary straight and parallel to the electron beam and continuous throughout the thickness of the specimen is small. Furthermore, the nickel and, in particular, the titanium surfaces will be partly oxidized when the specimen is exposed to air. Since diffusion along grain boundaries is much more rapid than bulk diffusion it is to be expected that the oxide surface layer extends farther into the specimen at the grain boundaries.

nanophase material one actually wants to characterize. Even in such favorable conditions the data obtained by HREM always have to be completed with more bulk-sensitive techniques. A method which is quite powerful, in particular in the characterization of the grain boundary structures, is total X-ray diffraction pattern analysis [84].

1.3.3 Precipitates

In most modern multicomponent materials, precipitates play an important part whether or not by intention. The nature of precipitation hardening, for example, aluminum with small amounts of other elements (e.g., magnesium, manganese, iron, copper, and silicon) is a classical example of process and materials understanding which could only progress in close relation with the development of the electron microscope [85].

Precipitates are introduced into aluminum, steel, and other metals and alloys to increase the strength of these. The nucleation, distribution, shape, and composition of the precipitation is of major importance to the performance of the composite [86]. In high-temperature superconductors, homogeneously distributed secondary phases can have a positive effect on flux pinning, thus increasing the critical current density [87]. In other materials the presence of impurities can lead to precipitation which may degrade the properties of, for example, semiconductor devices. Particles can be precipitated by chemical precipitation, temperature treatments (e.g., annealing of irradiated samples) or by mechanical working (defect-mediated solid state diffusion).

Microscopy techniques are used on precipitates to obtain information on phase identification, determination of the orientation relation between the precipitate and the matrix, and the relation between the size and morphology and the performance.

For phase identification and determination of size and morphology one can either separate the precipitates by dissolving the matrix, or thin the matrix and precipitates to electron transparency. The latter is done best by ion milling, although the precipitates can drop out from the specimen due to differences in ion milling rates; thus, a small angle of incidence of the ion beam is generally required. With a thin specimen of matrix plus precipitates the particles are in general more stable, but contamination is usually much greater. Phase identification can be done by electron diffraction, chemical analysis or HREM. If the particle is not continuous through the specimen, thickness difficulties arise, for example a background signal from the matrix in EDX analysis, Moiré interference fringes in high-resolution imaging, and unclear grain boundaries. Moiré fringes could also be used for identification of the precipitate and its orientation. Normally, either dark field techniques or elemental mapping are used for collective imaging of precipitates and determination of the distribution of precipitates.

The determination of the orientation relations between precipitates and matrix should be done by electron diffraction, either in SEM if the precipitates are sufficiently large or in TEM. If the precipitates are very small, a transmission electron microscope or scanning transmission electron microscope equipped with a field emission gun is essential. Similar difficulties occur as for the phase determination.

For TEM and STEM one uses an ion-milled specimen. For SEM, one needs a polished and clean surface. The effect of strain on the interaction between precipitates and matrix can be studied using strain holders [88].

1.3.4 Magnetic Materials

Evidently, magnetic materials are of enormous importance in the present methods for storage of information. An important development is the continuing miniaturization of magnetic particles, thus allowing a higher storage density. This is for a major part achieved by the fabrication of thin films. A comprehensive review of the state of the art of the fabrication and characterization of surface, interface and thin film magnetism is given by Falicov et al. [89]. The magnetic properties of fine particles are reviewed by Dorman et al. [90].

The microstructure can drastically affect the magnetic properties of a material. Micromagnetic properties such as hysteresis and the positions of domain walls depend critically on the presence of grain boundaries, free surfaces, defects, and voids. Properties pertaining to classical magnetism, such as magnetization and crystalline anisotropy, can also be substantially changed by irregularities in the atomic positions, in particular when the latter are different from the bulk positions due to the presence of a surface, a defect, or a grain boundary. Victoria [91] has used theoretical techniques to show the relation between magnetic behavior and microstructure.

By layering a magnetic material with a nonmagnetic material it is possible to produce thin films with unique properties, such as giant magnetoresistance [92], an enhanced magneto-optical Kerr effect [93], or a large perpendicular magnetic anisotropy [94]. Because of the large perpendicular magnetic anisotropy of ultrathin cobalt layers (about 0.4 nm) and the large polar Kerr effect, the Co–Pt multilayer [95] is one of the candidates for rewritable magneto-optical disks [96].

The properties of the multilayers are quite different from the properties of the bulk materials, due to the artificially fabricated microstructures. For instance, the giant magnetoresistance of Fe–Cr depends strongly on the thicknesses of the magnetic and nonmagnetic layers [97]. Since these properties are intimately related to the microstructure, a thorough description of all the microstructural features is required. Such a description includes the crystalline characteristics (e.g., grain sizes, orientations, orientation relationships, compositions, and compositional changes along the multilayer) and the magnetic microstructure (magnetization directions and domain walls).

Magnetic force microscopy (MFM) [98] is derived from AFM. It allows imaging of magnetic structures on a 50 nm scale. A review is given by Grütter et al. [99]. MFM has found applications in recording studies and the development of several types of recording materials [100].

Other techniques for observing magnetic microstructures include measurement of the magneto-optical Kerr effect [101], spin-polarized electron microscopy [102], electron holography [103], Lorentz microscopy, and the Bitter decoration technique [104].

Two specific problems in the electron microscopical characterization of magnetic particles occur because of the interaction

between the magnetic particles and the magnetic field of the objective lens. In the first place the magnetic field of the particles results in a distortion of the magnetic field of the objective lens. This can be easily adjusted with astigmators, but the adjustment has to be repeated after every change in specimen position. It is best to use as little magnetic material as possible. The second problem is that the particles or conglomerates of particles can be broken out from the specimen by the magnetic field of the objective lens which can lead to a 'permanent' contamination of the objective lens. If a high resolution is not needed, one can use a microscope with a switched-off objective lens. In particular with an extra post-column lens one can obtain magnifications up to about 60 000 times. Without the objective lens no high-resolution imaging is possible because the quality of all lenses except for the objective lens is not high enough.

Regarding multilayers, their magnetic properties depend strongly on structural details such as the texture, layer roughness, sharpness of the interfaces between the layers, and the presence of defects; consequently, these properties depend also on the synthesis conditions. Thus, a detailed characterization of the microstructure using HREM in relation to the synthesis conditions and the properties is required.

1.4 Required Characterizations

In this section the required characterizations as listed in Table 1 are discussed, with a strong emphasis on the use of TEM.

1.4.1 Size and Shape of the Particles

In the assessment of the size of small particles a whole range of microscopy and other techniques have been applied. For the very small sizes only SEM, TEM, STEM, and, partly, STM give a direct and individual image of the particles. Conventional X-ray and neutron powder diffraction only yield an indirect and averaged estimate of particle size through their size effect on line broadening. With these techniques, particles in agglomerates or part of twinned particles are detected as single particles, resulting in an overestimation of the available surface area. On the other hand, low-angle diffraction registers density variations such that agglomerates and single particles cannot be distinguished.

Intrinsic drawbacks in the use of STM in particle size analysis are that the resulting image is a convolution of the tip and the particle profile, and that one images only the surface of the particle; thus the shape at the side of the support cannot be detected.

If the size of isolated particles (for instance in a fluid) is needed, they have to be mounted on a support for TEM as well as SPM investigations. This mounting may change the morphology because of the particle–support interaction [105].

Inclusions and precipitates (not possible to image using STM) occur in their natural surroundings. Thus, determination of the size and shape of the particles is not hampered by changes provided the sample preparation does not lead to artefacts. However, one has to take into account that the TEM image is the two-dimensional projection of the

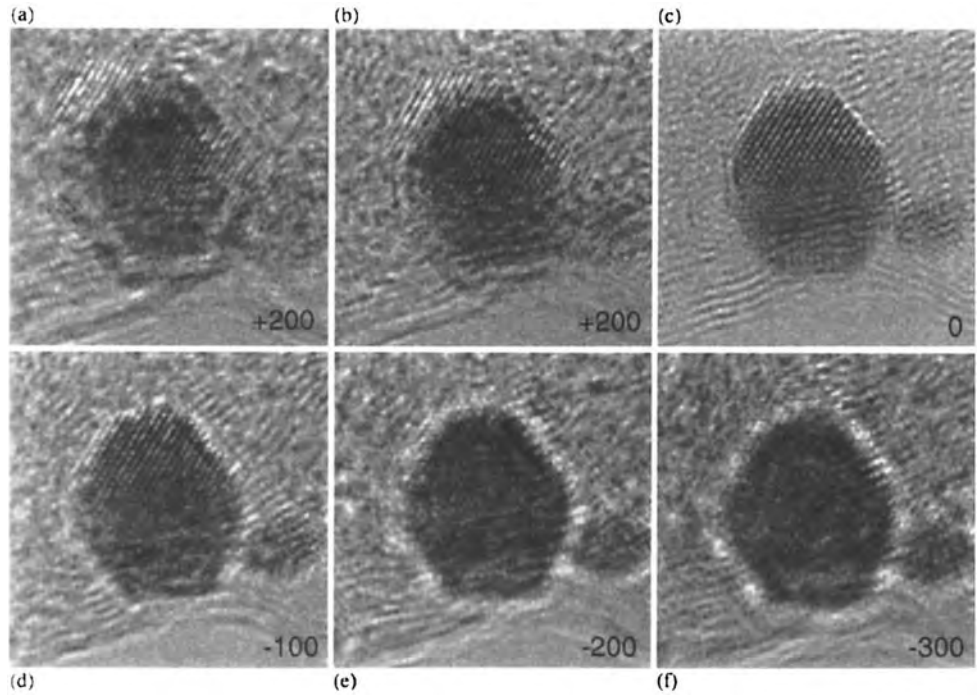


Figure 6. The effect of defocus on the size of a small particle. (a)–(f) A gold particle at focus values of about −200, −100, 0, 100, 200, and 300 nm, respectively. Note that the apparent size of the particle changes considerably and that at overfocus, in particular, the size cannot be determined.

three-dimensional shape of the particle. Thus, the apparent particle size (size distribution) in the TEM image can be affected by the shape and of the thin foil. It is then necessary to use some theoretical considerations to correlate the apparent size with the real size [106].

The TEM imaging conditions can strongly influence the apparent size of the particle. An example of this is given in Fig. 6, which shows a focus series in which the apparent size changes when the focus is changed. The observed discrepancy can be as large as 100% due to delocalization of information. Note that the size is especially difficult to estimate for an overfocus situation.

1.4.2 Structure of the Particles

1.4.2.1 Comparison of Transmission Electron Microscopy with X-Ray or Neutron Diffraction

The limited size of small particles has always brought about difficulties in their structural characterization. The line broadening in X-ray or neutron diffraction patterns due to the limited size of the particles has to be taken into account in the structural characterization, and thus knowledge about size and shape are required. The latter kind of data can be obtained by conventional electron microscopy, after which structure determination from X-ray or neutron powder diffraction

is more feasible. A more straightforward technique, however, is the use of HREM, which allows a direct structure determination, provided the specimen is sufficiently thin and images can be made in several low-index orientations. Information in diffraction space of the TEM goes down to more than $12\,nm^{-1}$, whereas information in HREM imaging depends on the quality of the objective lens and the energy spread of the electrons, and is about $7\,nm^{-1}$ for conventional high-resolution electron microscopes. Quite often one can verify with HREM whether the structure of the particle is the same as the bulk structure. An additional reason for the use of HREM is that many small metal particles tend to be multiply twinned. The structure of multiply twinned particles cannot be solved by conventional diffraction techniques if the twinning density is high.

If the basic structure of a metal particle is known and relatively simple (simple single-element metals), and if the twin boundary is simple and straight along the electron beam, one can deduce the structure of the twin boundary relatively easily from a conventional HREM image. In the case of more complex structures the deduction of the structure of the twin boundary can be very complicated. An important reason for this is the presence of delocalization, caused by the transfer function of the electron microscope. An example of this can be seen in Fig. 7, which shows an Si–Al precipitate in an aluminum matrix.

Evidently, structure analysis by HREM is local. On the one hand this allows structure determination of a minority phase, which is not possible by X-ray or neutron diffraction. On the other hand, the

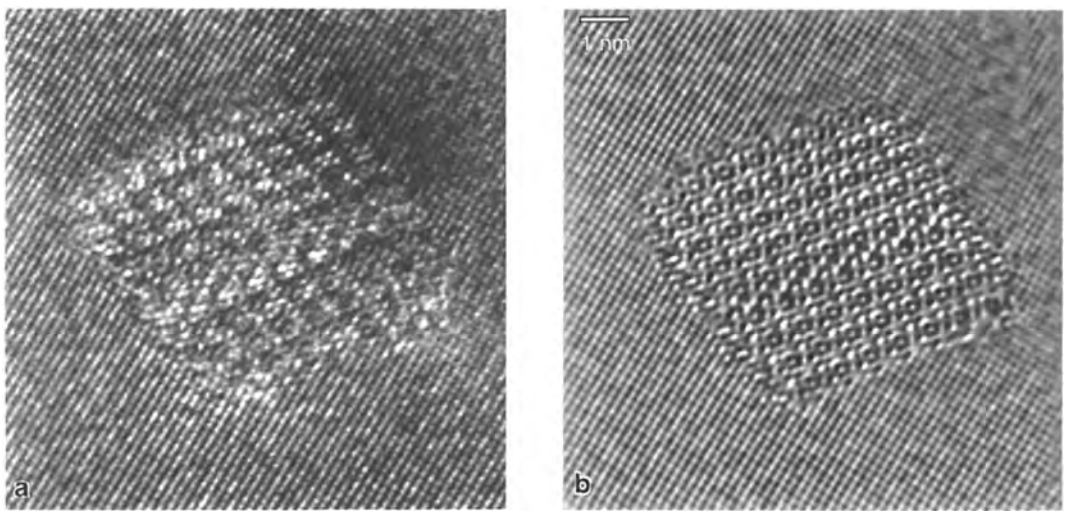


Figure 7. HREM images of an Mg–Si precipitate in an aluminum matrix, showing the effect of delocalization of information. The images were made with a Philips CM30ST-FEG. (a) An experimental image taken at a focus of approximately −70 nm. (b) The exit wave (phase part of the complex image), which is calculated from a series of 20 experimental images of which (a) is the second image. The exit wave shows sharp precipitate–substrate interfaces and a defect in the precipitate which is localized in one layer, whereas the interfaces in image (a) are vague and the defect seems to expand over about 1 nm. The less sharp image features in (a) are the result of delocalization of information.

specimen has to be very thin, which can lead to artefacts.

1.4.2.2 Use of the Electron Microscope

Unit Cell Determination

If one is interested only in the determination of the unit cell of a small particle, nanodiffraction using a spot size smaller than the particle is a very quick and easy method. For this, diffraction data are needed for several orientations of the particle. In principle these orientation could be chosen at random, as only the orientation relationships between the electron diffraction patterns should be known. However, it is easier to orient the small particle in a low-index zone and subsequently tilt it about one of the (small) reciprocal lattice vectors.

Structure Determination Using Kinematic Diffraction Theory

One can also determine the structure from the electron diffraction data by a kinematic approach is the diffracting area is very thin (<3 nm). This small thickness is essential because the interaction of the electrons with a specimen is very strong, resulting in dynamic scattering, which changes the intensities of all reflections with respect to each other as a function of the specimen thickness. If the dynamic diffraction is large, the intensities of the reflections cannot be used for structure determination unless the dynamic scattering is taken into account. Provided the crystal is very thin (2–4 nm), a structure refinement based on the kinematic diffraction theory yields acceptable results. In this case one can use the structure refinement programs, which are well developed for X-ray diffraction and neutron diffraction. Refinement of simulated electron diffraction data shows that kinematic refinement of electron diffraction data requires the specimen thickness to be in the range of 1–5 nm, depending on the scattering potential of the specimen. The advantage of a kinematic refinement is that the standard structure refinement techniques used in X-ray single-crystal diffraction can be used.

Structure Determination Using Dynamic Diffraction Theory

Four techniques have been reported in which dynamic diffraction is taken into account in the refinement process:

(i) Zuo and Spence [107] use convergent beam diffraction to refine structure factors. From these, structural parameters can be deduced. The method is limited to small unit cells.

(ii) Tsuda and Tanaka were able to refine a small number of positional and thermal parameters of $SrTiO_3$ using convergent beam patterns [108]. It proved to work for very thick crystals for which the thickness can be estimated with convergent beam techniques. The usefulness of this method is limited because of the large calculation time. Since many CBED patterns have to be calculated, only a very limited number of parameters can be taken into account. More parameters will increase the computing time exponentially. All current Debye–Waller factors and positional parameters were reported to be refined.

(iii) Sha et al. [109] propose a different approach using normal diffraction patterns. They correct for dynamic

Table 3. Atomic positions obtained by refinement of several sets of [0 0 1] electron diffraction (ED) patterns of $Ce_5Cu_{12}P_{19}$.

Atom	(position)	x (XSD)	x (ED)	y (XSD)	y (ED)
Ce1	$(\frac{2}{3}, \frac{1}{3}, 0)$				
Ce2	$(x, 0, \frac{1}{2})$	0.8067(1)	0.813(2)		
Cu1	$(0, 0, 0)$	0.2878(2)	0.287(2)		
Cu2	$(x, 0, \frac{1}{2})$				
Cu3	$(x, y, 0)$	0.3783(2)	0.378(1)	0.1725(2)	0.174(2)
Cu4	$(x, y, \frac{1}{2})$	0.6359(2)	0.636(2)	0.1185(2)	0.117(2)
Cu5	$(x, 0, 0)$	0.4508(4)	0.445(2)		
P1	$(x, 0, 0)$	0.1768(5)	0.178(2)		
P2	$(x, 0, 0)$	0.6298(5)	0.641(2)		
P3	$(x, y, \frac{1}{2})$	0.3216(4)	0.319(1)	0.8454(4)	0.830(1)

$Ce_5Cu_{12}P_{19}$ has a hexagonal space group P-632, with $a = 1.24$ nm and $c = 0.40$ nm. Crystal thicknesses were refined, and ranged from 12 to 15 nm. For comparison, the atomic positions for $La_5Cu_{12}P_{19}$, obtained from X-ray single-crystal diffraction (XSD) are given. The R value was 5.2%.

diffraction based on a rough atomic model to create a resulting quasi-kinematic data set, which can subsequently be used in a standard kinematic refinement procedure.

(iv) Jansen et al. [110] have developed a computer program, multi slice least square fit (MSLS), which incorporates the dynamic diffraction theory into the structure refinement process. The program combines two known algorithms: the multislice method and a least-squares fitting procedure. The calculation takes into account the curvature of the Ewald sphere and multiple scattering. Using this method, one can use much thicker crystals (e.g., 14 nm for $Ce_5Cu_{12}P_{19}$) [111] than can be used for a kinematic refinement (see Table 3).

The diffraction data, used for the refinement of $Ce_5Cu_{12}P_{19}$, were taken from areas with a diameter of about 10 nm. Thus, this procedure allows a detailed structure determination of particles with sizes from 5 nm and larger, which makes it a quite powerful method for structure determination of small particles; it also gives a figure of merit for the refined model.

1.4.3 Composition (Overall) of Particles

The composition can be determined with the well-developed techniques of energy dispersive X-ray analysis (EDX), wavelength dispersive X-ray analysis (WDX), and EELS. In the case of EDX and/or WDX the technique is generally called electron probe microanalysis (EPMA). Evidently, the technique to be applied depends on the particle size and what information and accuracy is needed. A restriction in the use of element analysis of small particles can be the broadening of the electron beam due to the electron–specimen interaction. For instance, using

100 keV electrons the beam broadening is 1 nm for 30 nm thick γ-Al_2O_3 (and 20 nm for a thickness of 200 nm) [112].

EPMA is an accessible and often used method for determining the element distribution in micrometer-sized particles. For smaller particles, however, the difficulties with quantitative EPMA (absorption, geometry, fluorescence, and atomic number) experienced in ordinary samples are accentuated [113]. These difficulties are related to beam broadening, resulting in a strong loss of resolution. The resolution limit can be improved by changing to a lower electron acceleration voltage, or by using thin sections (similar to TEM) of embedded (if not already in a matrix) particles. Thin sections can be obtained by, for example, ultramicrotomy.

A problem in the composition determination using X-rays in TEM is spurious radiation, resulting in X-ray generation outside the area illuminated with the electron beam. This radiation can be due to X-rays generated by stray electrons somewhere in the microscope column and not stopped by apertures, or by secondary electrons created in the illuminated area of the specimen. Evidently the first cause can be reduced by the design of the microscope, but the latter is inherent to the technique.

In many systems it is only necessary to detect or confirm that a certain element is present, or to locate precipitates and inclusions within a matrix or grain. This can be done by elemental mapping using EDX and/or EELS in a scanning transmission electron microscope or in transmission electron microscope (in nanoprobe/scanning mode) or by electron spectroscopic imaging using energy filtering [114]. When the individual composition of each particle is required, a widespread technique is quantitative X-ray analysis using the microscope response factors for the various elements [115]. For many simple particle systems, such as single-element particles or bi-element particles, this is the fastest and easiest method. It becomes increasingly difficult to quantify reliably the light elements due to absorption effects. It also should be taken into account that the reproducibility of a window-less detector is lost if contamination has accumulated on the detector during the period of investigation. For small and/or light element particles, EELS may be advantageous due to its better sensitivity to light elements, sensitivity to chemical bonding, and higher collection efficiency. With more complex multi-element systems and compounds containing light and heavy elements (e.g., oxides), a combination of several techniques should be used. Lyman et al. [116] observed a correlation between the composition and the size of supported metal catalysts, in which it is sufficient to measure the particle size. If the composition is strictly related to a structure (e.g., some metal oxides or inclusions) one can also perform microdiffraction to distinguish a certain particle.

Problems in TEM and STEM that have to be faced in quantitative analysis of nanometer-sized particles are drift, the trade-off between a high count rate (good statistics) and the high electron intensity that the particle can sustain (decomposition). Since the smaller the particle to be analysed the larger the effect of specimen drift, one needs for particles smaller than 5 nm a high electron beam intensity, which can be readily obtained with microscopes equipped with a field emission gun.

1.4.4 Structure of the Surface

1.4.4.1 Polarity of the Surface

In the case of small particles of oxides or other partly ionic compounds, the interruption of the lattice at the surface must lead to some changes of the surface layer to obtain charge compensation. The simplest situation occurs if the ending layer is charge-neutral. In that case the layer is stable in itself, and, by a small relaxation of the outer atoms, the effect of the broken bonds can be reduced. If the ending layer is not neutral it will have a high surface energy, which makes it unlikely that such a surface plane can exist. However, different atoms can be physisorbed or chemisorbed (e.g., OH^- groups on an MgO surface), which can result in a reduction of the surface energy such that the surface becomes more stable. Consequently, the atmosphere in which the particles are formed in crucial in the morphology of these small particles.

1.4.4.2 Imaging Surface Relaxation by Transmission Electron Microscopy Profile Imaging

The extent of surface contraction or repulsion has been studied by Marks [117], who has done extensive calculations on this subject. He showed that the apparent expansion at the surface, as measured from a high-resolution electron micrograph, depends on the defocus of the electron microscope. Thus the observed positions of the surface atoms are only equal to the actual atom positions if the correct focus is chosen. He also indicated that when one takes into account the noise level of the images, the real position of the atomic columns could be determined with an accuracy of about 5% of the lattice parameter.

The problem of artefacts due to the focus of the transmission electron microscope can be partly overcome by the use of exit wave reconstructions (see Chap. IV, Sec. 1.1.2 of Vol. 1 of this handbook). As in conventional HREM, the absolute focus is not known in the exit wave reconstruction. This requires a method of finding the absolute focus. Tang et al. have shown that this can be done after the reconstruction of the absolute focus using a maximum entropy method [118], yielding the absolute focus with a precision better than 5 nm. Figure 8 shows the effect of +5, 0 and −5 nm focus shifts from the absolute focus for an exit wave of a (1 0 0) surface of NiO. The focus shifts of +5 and −5 nm result in shifts of −0.02 and +0.02 away from the bulk, respectively.

The atomic structure of steps, adatoms, vacancies, and other surface defects can be readily investigated with STM [119]. A lateral resolution of 1 Å and a vertical resolution of 0.1 Å are achieved routinely.

1.4.4.3 Can One Detect Ad-Atoms?

It is not likely that one is able to image adatoms by HREM with a detail that allows one to draw conclusions. Probably the best HREM configuration is to use profile imaging. However, to image a single adatom next to columns of atoms of the surface layer is difficult. This becomes

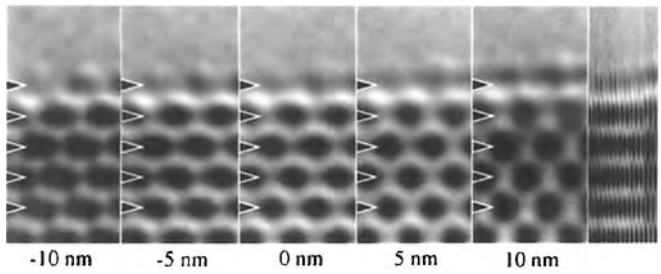


Figure 8. The effect of a focus propagation on profile images of the (0 0 1) surface of NiO. The applied focus propagations are given below the images. The image on the far right represents the same series of images but compressed in the horizontal direction, which is approximately the same as looking at a glancing angle in the horizontal direction. The position of the row of dots at the surface is shifted away from the bulk when a negative focus propagator is applied.

even more difficult if the ad-atom has a large Debye–Waller factor, which is highly likely to occur, because it has relatively few bonds with the substrate. A high Debye–Waller factor results in a further decrease in the contrast of the ad-atom. The effect of a high Debye–Waller factor is illustrated in Fig. 9, showing exit wave profile images of the (0 0 1) surface of NiO. The dots of the nickel columns in the amplitude at the surface are less dark, which is due to the larger thermal vibration (and not to fewer atoms in the nickel column), which is quite normal for atoms at the surface.

For the imaging of ad-atoms, SPM is a better technique. Many illustrations of the unique possibilities of this technique have already been reported [120]. For instance, STM on a closed ad-layer of sulfur on a rhenium (0 0 1) surface is reported by Ogletree et al. [121] to form a $2\sqrt{3} \times 2\sqrt{3})R30^\circ$ superstructure, whereas CO on a ruthenium (0 0 1) surface forms 2×2 superstructure [122]. Individual ad-atoms are, for instance, observed for carbon on aluminum (1 1 1) [123]. Examples of processes at the metal–gas interface are changes in the oxygen coverage of ruthenium (0 0 1) [124] and the movement of oxygen strings on copper (110) [125].

1.4.5 Composition of Particle Surfaces

The accuracy of TEM or STEM for the determination of the composition of the surface of small particles is much less than that of surface-sensitive techniques such as SIMS and XPS. Thus, it is not useful to apply TEM or STEM for this purpose; also, it is often desired to probe the surface as a function of temperature or exposure to certain gases, which cannot be done in electron microscopes. STM could be used, but one needs very precise knowledge on the density of states of the surface atoms to allow interpretation of the images (see Sec. 1.5.2 of this Chapter).

If the surface composition can be associated with a specific structure feature by HREM, then the surface composition can be inferred from atomic profile imaging of the particle surface, for example ZnO on $ZnCrFeO_4$ [126] or TiO on TiO_2 [127].

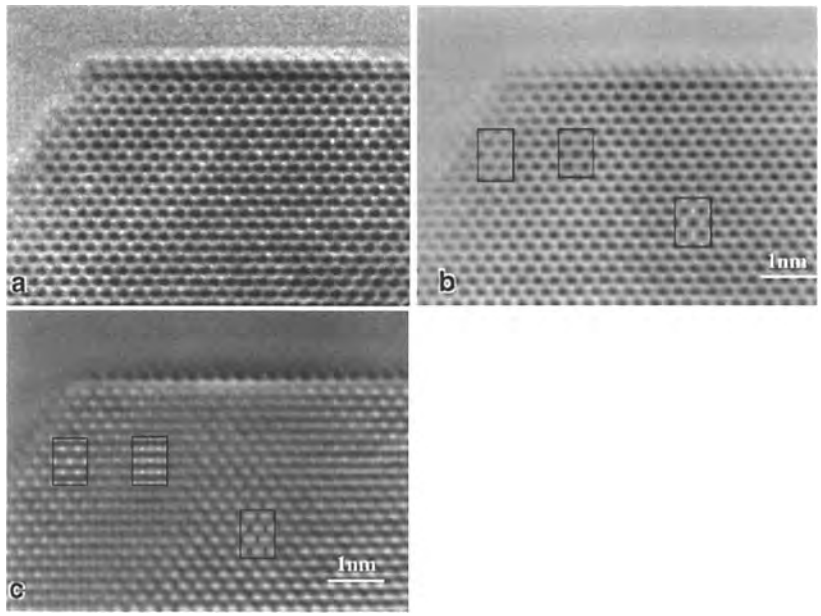


Figure 9. Profile images of the (0 0 1) surface of NiO. (a) An experimental image taken at −100 nm with a Philips CM30ST. (b,c) The amplitude and phase of the exit wave, calculated from a through-focus series of 15 images, of which image (a) is the first image. The black dots correspond to nickel columns in the viewing direction. The nickel atoms are imaged as black dots in the amplitude, whereas the white dots in the phase image correspond to the nickel positions in the thin part of the crystal and to oxygen positions in the thicker part. The black dots of the nickel columns in the amplitude at the surface are less dark, due to the larger thermal vibration, which is quite normal for atoms at the surface. The experimental image does not show this feature of less contrast at the surface because of delocalization and because it contains information from both the amplitude and the phase of the exit wave (the phase is less sensitive to differences in thermal vibration).

1.4.6 Structure of Grain Boundaries and Orientation Relationships of Adjacent Grains

The determination of grain boundary structures requires another approach to the determination of the orientation relationships between adjacent grains. Two important questions in the study of grain boundaries to consider are the structure of the grain boundary and the orientation of the grains.

The determination of the orientation relationships between adjacent grains of micrometer-sized particles can be done using SEM in either of the two modes, giving either an electron back-scatter pattern [128] or a selected-area channeling pattern [129]. For submicrometer particles, the orientation can be revealed by taking TEM diffraction patterns of each grain, or microdiffraction patterns if the grains are small. For ensembles of rather low-angle grain boundaries, methods have been developed based on obtaining Kikuchi patterns for each individual grain [130]. With the help of a computer this method is relatively fast and precise.

In many cases, sufficient knowledge of the grain boundary can be inferred from knowing the orientation of the adjacent crystals. Then the task reduces to a determination of crystal orientations. The only

way to obtain direct information of the grain boundary structure, for example how the atoms are arranged or whether there is an amorphous interlayer, is to reveal it using HREM. This, however, is far from trivial as many experimental conditions have to be satisfied:

(1) the boundary plane needs to be (almost) parallel to the electron beam;
(2) one low-order zone axis of both crystals needs to be aligned with the beam; and
(3) not too much strain should be present.

These conditions are usually only fulfilled for low-angle grain boundaries or special grain boundaries (e.g., pure tilt, symmetrical). The chance of getting close to fulfilling the stated conditions is increased for extremely thin specimens. However, other difficulties do occur with such ultrathin specimens: (i) it is difficult to prepare a very thin specimen with equal thickness over the area of interest, in particular because etching rates are orientation-dependent, and (ii) the structure of the grain boundary can be changed due to exposure to air or due to relaxation.

1.4.7 Composition of Particle Grain Boundaries

As in the composition analysis of surfaces the analysis of the composition of a grain boundary by TEM techniques is very difficult. Additional to the problems with the surface analysis, it is necessary to consider the orientation of the boundary plane with respect to the plane spanned by the detector and the electron beam, because this can influence the accuracy (or at least the spatial accuracy) of a quantitative analysis. Surface-sensitive techniques can be applied only if the material can be fractured at the grain boundaries.

1.4.8 Ageing

The performance of most materials changes in time, which is called ageing in catalyst research. To establish the mechanisms involved in ageing, a sequence of samples can be taken as ageing progresses. From the characterization of these samples, it is possible to determine the mechanisms, but in an indirect way. A more direct way is to perform the ageing on a region which is constantly monitored by a microscopic technique or to investigate the same area at certain moments of the ageing process. The first approach is restricted by the conditions required for performing the microscopic technique (TEM—vacuum, electron transparent specimen; STM—only the surface can be imaged). In this respect one can use an in situ cell (see Sec. 1.8.5 of this Chapter) or special TEM specimen holders, which allow cooling, heating, or applying strain to the specimen. Regarding the second approach, one needs a vacuum transfer system from the ageing equipment into the electron microscope and vice versa (see Sec. 1.8.6 of this Chapter).

1.5 Detailed Structure Imaging

1.5.1 Delocalization of Information

A significant problem in the characterization of small particles by HREM is the

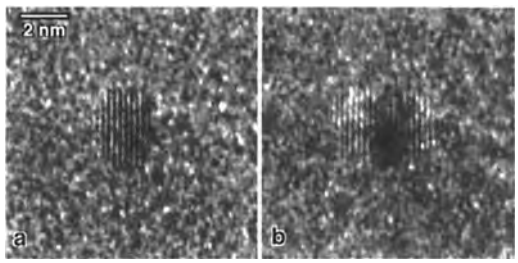


Figure 10. Delocalization of (1 1 1) fringes of platinum particles on a silica substrate. (a) The particle at about Scherzer focus (−60 nm). The (1 1 1) fringes are located close to the actual position of the particle. (b) The same particle but at a focus of −300 nm. The (1 1 1) and ( ) fringes are shifted in a direction parallel to the corresponding $\boldsymbol{g}$ vectors. The shift of these fringes is equal to gel where e and l are the defocus change and the wavelength respectively.

delocalization of information. Delocalization is illustrated in Fig. 10, which shows a platinum particle at approximately −60 and −300 nm defocus. The image at −60 nm (about Scherzer focus) shows the platinum particle with its (1 1 1) fringes at the same position as the dark feature due to the amplitude contrast of the particle. With changing focus the (1 1 1) fringes shift, resulting in a shift of about 1.5 nm for a defocus of −300 nm. The (1 1 1) and ( ) fringes are shifted in opposite directions. The shift of these fringes is equal to $\boldsymbol{g} \times \Delta f \times \lambda$ where $\boldsymbol{g}$, Δf, and λ are the reciprocal lattice vector, the defocus change, and the wavelength, respectively. Since most frequencies describing the shape of the diffraction contrast image of the particle are small, the shape is still easily visible in Fig. 10b, and is located at the actual position of the particle.

The effect of the delocalization on grain boundaries is illustrated in Fig. 7, showing an Mg–Si precipitate in an aluminum matrix. The figure shows an experimental image taken at −70 nm with a Philips CM30ST and an exit wave, calculated from a through-focus series of 20 images of which the image at −70 nm defocus is the second image. The Mg–Si precipitate contains a defect. There are three differences between the two images. The first difference is that the single image is much noisier, which is logical because for the exit wave the information from 20 images is taken, resulting in a reduction of the random noise. The second difference is that the defect in the Mg–Si precipitate is quite sharply imaged in the exit wave whereas in the experimental image the defect region seems to be much broader, which is the result of the delocalization. The third difference is that the image of the periodic structure of the precipitate is much clearer in the exit wave image, which is due to a correction for the delocalization, as for the grain boundaries.

Another advantage of the use of through-focus exit wave reconstruction is that one can use the phase and amplitude information in combination. Figure 11 shows an example of a MoS_2 slab on g-Al_2O_3. An experimental image and the exit wave reconstructed from a series of 15 images is shown. Whereas the phase image shows dark dots at the molybdenum and sulfur columns, the amplitude image highlights the molybdenum columns.

1.5.2 Image Calculations

In TEM there exist two types of image simulations: HREM (phase contrast) simulation and diffraction contrast calculation (amplitude contrast). The ultimate aim is to relate simulated contrast quantitatively to the experimental contrast, and in this way to extract important

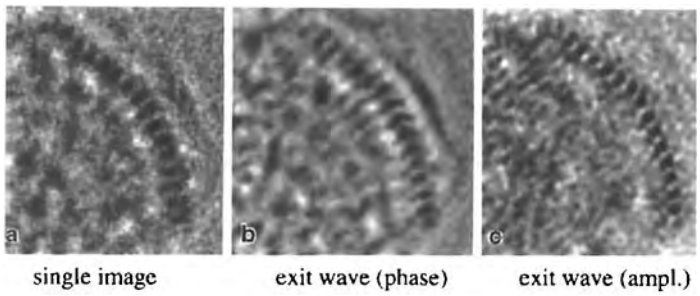


Figure 11. An experimental HREM image and an exit wave (phase and amplitude) of an MoS_2 slab on g-Al_2O_3. The exit wave is reconstructed from a through-focus series of 15 HREM images, of which the experimental image is the first. Whereas the phase of the exit wave shows dark dots at the molybdenum and sulfur columns, the amplitude of the exit wave highlights only the molybdenum atoms.

parameters about a particle. A general problem is to include inelastic scattering or electron beam-induced changes.

Simulation of HREM images of small particles can be done with most existing software packages. The unit cell needed to do a calculation is typically the whole particle, for example a spherical copper cluster with a diameter of 3 nm is roughly 1100 atoms. Experiments and HREM contrast simulation have been performed on clusters/clouds of, for example, gold and carbon on a gold support [131]. The calculations become more complex if an unknown particle–support interaction has to be taken into account. Giorgio et al. have done experiments and simulations on a Pd–MgO system where a lattice distortion due to the substrate interaction was found for 1–2 nm particles [132]. In the selection of models for such a particle–support interaction it may be helpful to perform molecular dynamics calculations (or similar) rather than using ad hoc constructed models.

Diffraction contrast calculation can be used to show not only the strain field around defects and stacking faults [133] but also around coherent precipitates, and the adhesion strain field in small particles [134]. Extinction rings in particles down to 10 nm can be observed in bright field and can be compared to simulated images [135]. Extinction rings can even be observed in smaller particles using weak beam dark field techniques [136]. Thus, adhesion strain fields allow extraction of such information as surface energy or the chemical fingerprint of a particle. In the process of proposing and calculating models of strain fields of defects, precipitates, and adhesion, it may prove advantageous to do molecular dynamics calculations as well. Since the latter calculations are able to handle more than 10^6 atoms they allow calculations of defects with a volume as large as $25 \times 25 \times 25\,nm^3$ [137].

Whereas by STM the atomic structure of steps, ad-atoms, vacancies, and other surface defects can be readily investigated [138], the interpretation of these images is certainly not straightforward. It is necessary to perform image calculations, for which one needs the local densities of states. Software to do total energy calculations of surfaces has been developed for metal surfaces, but is still under development for more complicated systems such as oxide surfaces or metal particles on an oxide substrate (one could say that image

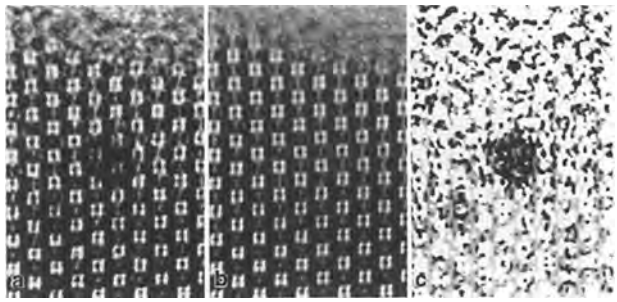


Figure 12. Image processing on an HREM image of a platinum particle in zeolite ZSM-5. (a) The raw HREM image. (b) The image after patching the area next to the platinum particle on the area of the platinum particle and averaging over four unit cells (2 × 2). (c) The difference between (a) and (b). (c) It can be seen that the platinum particle is circular, indicating a spherical particle. This particle is too large (about 2 nm) to fit into the pores of ZSM-5 (which are about 0.7 nm).

calculation for SPM is now at the level where image calculation on HREM images was about 10 years ago).

1.5.3 Averaging of Images

Averaging of images can be done in time or in space. In the latter case one averages several copies of the same image which are displaced over n times a lattice vector. In general, this is done if the information required is periodic over the vectors over which it is averaged, giving an image in which noise and local changes, which are not related to the required information, are reduced. Averaging in time is, of course, done for a given exposure time. Because of specimen drift, the exposure time has to be limited but a series of images can be taken, which by cross-correlation can be added, thus increasing the exposure time markedly [139]. Both methods are particularly useful in the investigation of materials which are beam-sensitive. An example of averaging in space is given in Fig. 12, which shows an image of Pt-loaded ZSM-5 [140]. In this example the image of the zeolite is averaged, giving an average lattice image; next, this averaged lattice image is subtracted from an image of the zeolite with a platinum particle, yielding an image from which the shape of the platinum particle can be much better determined.

1.6 Beam-Induced Phenomena

Electron beam-induced changes are mostly either a function of electron dose or electron flux. Thus, one can reduce these changes by working fast or using a lower electron flux, respectively. If one starts with a new type of material, it is advisable to record also the information (images or electron diffraction patterns) as a function of beam current. An example of electron beam-stimulated decomposition is shown in Fig. 13. This figure shows a series of HREM images of the decomposition of a single WS_2 slab into tungsten. It was found that this decomposition is strongly enhanced if the specimen is exposed to air, which indicates that adsorbed molecules

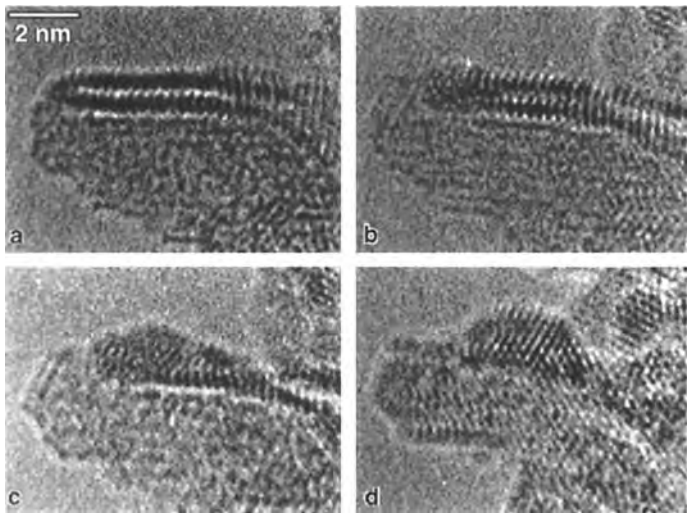


Figure 13. Series of HREM images from the same area recorded with time intervals of 2 min. The series shows the decomposition of a single WS_2 slab into tungsten. The decomposition starts at the edges of the WS_2 slab, indicating this area to have the lowest stability.

from the air (H_2O, OH^-) are an important factor in the decomposition. This points to the need for better specimen preparation procedures.

Apart from the decomposition of a phase, the surface of a specimen can be etched/evaporated by an intense electron beam. For instance, 100 keV electrons have been observed to etch holes in thin foils of crystalline and amorphous aluminum [141], creating voids and holes on a nanometer scale (1 nm in diameter) parallel to the electron beam. MgO molecules/clusters were observed to be desorbed from MgO thin films, producing square indentations and holes [142]. Similar square indentations/holes have been observed by the author to appear in $SrTiO_3$ particles under heavy electron radiation. Small particles are basically subject to similar extreme conditions when irradiated by a high-energy electron beam. Mass loss due to electron irradiation is probably closely related to particle temperature and the build-up of an electric field (charging, Auger process), possibly both in combination. Many indications have been reported that particles on supports are as susceptible to radiation damage in the sense of electron-stimulated structure alteration. There have been frequent observations of, for example, the movement and growth of metal clusters, atomic displacement, and crystal orientation changes [143]. Suddenly disappearing particles and particles vibrating violently are examples of charging effects.

Electron beam-stimulated structure fluctuations and atom desorption was suggested to be picoseconds temperature jumps above the melting temperature for gold clusters of about 1000 atoms. The mechanism reponsible for the increase in temperature was explained by M-shell ionization followed by an Auger relaxation/cascade [144]. Hobbs [145] and Reimer [146] give a general treatment of the energy deposition from an electron beam. Roughly two lines of explanation for the structural fluctations emerge [147]: (i) that a melting–recrystallization process takes place or (ii) that there is more than one 'ground state' structure that the particle can adopt when minimizing the energy.

A thorough treatment of the models and literature is given by Marks [148], alongside of which is suggested a model with a so-called 'quasi-molten' state which basically is linked to the broadening of the first-order transition by finite-size effects as found by molecular dynamics calculations [149]. The often observed difference between the beam intensity needed to start structural fluctuations and to sustain it can simply be related to the adhesive energy between particle and substrate that has to be overcome. The Auger process is further treated in detail by Cazaux [150], who emphasizes it as a major driving force for atom displacement and desorption especially in electrically insulating compounds. Extensive reference is given to example materials such as NaCl, SiO_2, Al_2O_3, MgO, and TiO_2.

1.6.1 Changes in Orientation of Particles and Particle Drifts

Some metal particles change their orientation almost continuously. Evidently, in these cases the possibility of imaging lattice fringes depends on the timescale of the orientation variations. They can be so rapid that one will not or only scarcely see some lattice fringes on a real-time camera. Quite often the rate of this change can be strongly increased or decreased by increasing or reducing the intensity of the electron beam, respectively. The most important parameter in this respect is the number of electrons per square nanometer, but quite often the total beam current on the specimen is important as well (due to heating of the surrounding area). Figure 14 shows a series of three images taken with time intervals of 2 s of gold particles on a carbon film. Due to the movement of the small metal particle towards the small one, a collapse into one particle occurs. As in this example, many other small particles also show an incorporation into a larger particle. Quite often one observes a slow movement of the particles and suddenly a very rapid movement (over 1–2 nm) of the small particle towards the larger one, such that they touch each other. Next, a sintering into one particle can occur, but one also

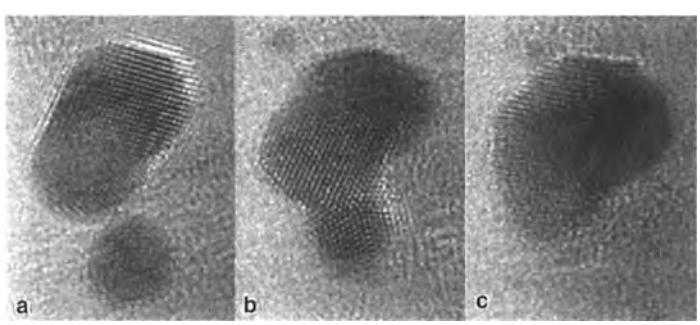


Figure 14. An example of orientation changes and sintering of gold particles, due to an intense electron beam which was about 20 times more intense than that used for normal recording of HREM images. (a)–(c) Images taken at time intervals of 3 s. In (a) two separate particles can be observed. In (b) the particles are connected, whereas they are just one particle in (c). Apart from this sintering to a larger particle, one can also observe the orientation changes of the larger particle.

observes quite often that the small particle is fixed to the larger one without sintering. This could be caused by an electrostatic repulsion or a surface layer on the particles.

1.6.2 Charging of Particles

Often catalyst/support specimens do charge. In particular, support bases on amorphous SiO_2 are susceptible to charging. Charging leads to vibration of the image and to an apparent objective astigmatism. If one is not interested in the surface of the catalytic particles, the charging can be reduced by coating the complete specimen (supported catalyst on the carbon-coated holey film) with a thin carbon layer (1–2 nm thick) (see also Sec. 1.6.5 of this Chapter). In our experience, charging can be mostly prevented by taking a few precautions:

(i) taking only small particles or ensembles which are well separated from larger particles or ensembles;
(ii) using a small probe size; and
(iii) selecting only those particles which are well supported by the carbon grid.

1.6.3 Contamination

The rate of contamination can vary strongly from one specimen to another. If that is the case, the contamination is not so much caused by the microscope but rather by the specimen itself. For instance, we have investigated a MoS_2–Al_2O_3 specimen in which with a relatively small beam (diameter 10 nm) no contamination was observed on MoS_2–Al_2O_3 particles that were over a hole in the carbon holey film, whereas a rapid contamination was observed when the beam was slightly shifted to an area where both MoS_2–Al_2O_3 particles and the supporting carbon film were irradiated.

1.6.4 Amorphization

Several catalytic materials (supports as well as the active phase) are susceptible to amorphization by the microscope. This process can be the result of the knock out of atoms (e.g., the amorphization of $YBa_2Cu_3O_7$ when 800 kV electrons are used) or of ionization. The latter is the process which leads to the amorphization of zeolites. For instance, a sodium-containing zeolite loses almost all its sodium by a reduction in Na^+ ions. Since these atoms stabilize the structure of the zeolite, to neutralize the partial substitution of Si^{4+} by Al^{3+}, it is logical that the removal or displacement of these atoms results in a collapse of the structure. Obviously the zeolite lattice is generally more stable for higher Si/Al ratios. The stability of zeolites can be improved by ion exchange with Li^+ or UO_2^{2+} ions [151].

The amorphization of zeolites can also be used in a positive way. In principle, metal particles inside the zeolite lattice can be detected, if the image is not too disrupted by the substrate. Obviously this requires that the zeolite is thin. However, the lattice of the zeolite must also not dominate. The effect of the zeolite lattice is shown in Fig. 15. If the zeolite is in a low-index orientation, most metal particles cannot be detected, no matter how

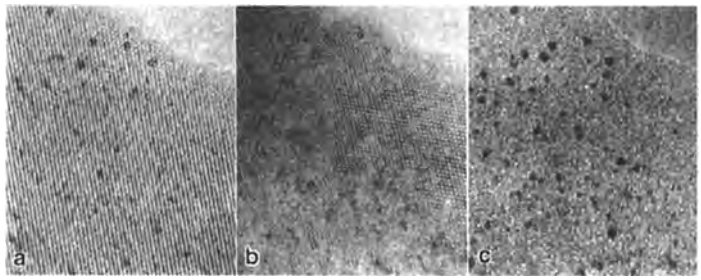


Figure 15. Platinum particles in a ZSM-5 zeolite lattice. (a) The zeolite tilted a few degrees from the [1 0 0] orientation. A number of platinum particles can be observed. (b) Image recorded after (a) and taken with the crystal in the [1 0 0] orientation. Hardly any platinum particles can be observed, which is due to the fact that the zeolite lattice is very dominant in this orientation. (c) The same area but after the zeolite has been amorphized by an intense electron beam. Platinum particles in the range from 0.7 to 2 nm can be clearly observed.

thin the crystal is. One can improve the visibility considerably by tilting the crystal out of the low-index orientation. But the visibility is greatly improved by amorphization of the zeolite lattice. An obvious question is whether the amorphization has led to sintering of the metal particles. This can be determined by comparison of the images with the zeolite lattice intact but in an orientation rotated from a low-index zone axis with an image with the same crystal orientation but after amorphization, for example Fig. 15.

1.6.5 Reduction of Surfaces

As is discussed below, the intense electron beam can cause radiolysis or even electron-stimulated desorption of atom species. This may be observed as complete depletion of a particle surface or as an alteration in the surface structure. Many observations have been reported of both oxidation and reduction under a 100–400 kV electron beam [152]. Figure 16a shows an example of the reduction of TiO_2 after

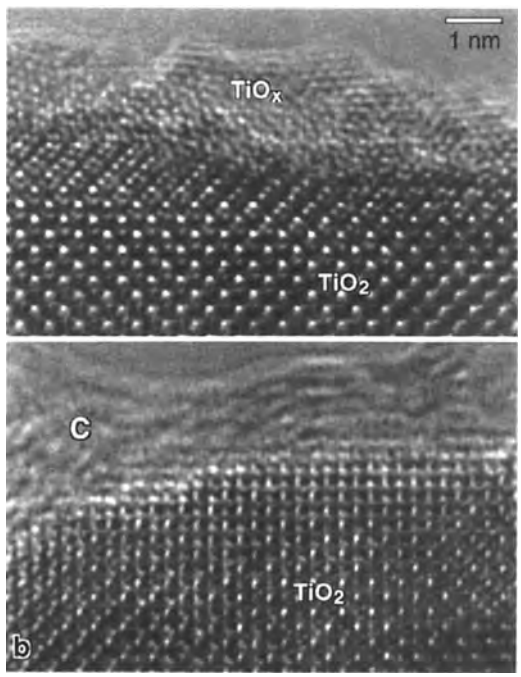


Figure 16. Example of the surface reduction of TiO_2. (a) A TiO_2 grain in the [0 0 1] orientation with a deviating structure at the edge, which is the result of the reduction of TiO_2 into TiO. This reduction progresses from the surface into the bulk upon continuing irradiation by the electron beam. (b) When TiO_2 surface is covered with a thin carbon layer, such a reduction does not occur.

10 min of exposure to an electron beam with an intensity used for HREM imaging. This reduction starts immediately and progresses steadily into the TiO_2 grain. Interestingly, however, in a similar experiment a thin carbon layer is apparently sufficient to protect the surface against radiolysis, as seen in Fig. 16b. Here the lattice image of the TiO_2 lattice parameter changes abruptly into that of amorphous carbon.

1.7 Typical Characterization Problems Related to Small Size

1.7.1 Effect of Substrate

The study of very small catalyst particles in the range of 0.3 to 1.0 nm is handicapped by the support on which the particles are located. Changes in the thickness of the support or rearrangements of the atoms at the surface of the support can lead to contrast features, which are quite similar to the contrast of the catalyst particles. In a number of cases one can do HREM to image the lattice fringes of the small particles, but quite often such catalyst particles are too small to distinguish clear lattice fringes. An example image of very small particles is shown in Fig. 4. These small particles are seen best when they are situated on very thin supports or when the surface of the substrate is thin and parallel to the electron beam. In all cases one should select those areas in which the support is very thin and, if crystalline, in an orientation in which no strong scattering occurs. Small particles can be seen best in profile imaging of very thin supports. An example of the effect of the blurring of the lattice image of a small particle by a crystalline support is shown in Fig. 17.

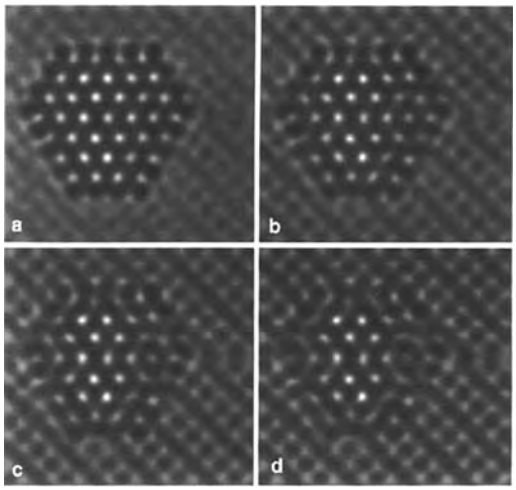


Figure 17. Image calculation of MoS_2 on a crystalline g-Al_2O_3 support. Images calculated for a single slab of MoS_2 with a thickness of 0.6 nm on top of the substrate with thicknesses of 1, 2, 3, and 4 nm are shown. The hexagonal MoS_2 lattice is only clearly visible for substrate thicknesses of 1 and 2 nm, in particular when random noise is added to the image.

In the case where one cannot observe any image features, owing to small particles, it is useful to do either local element analysis or to anneal this area with an intense electron beam, which generally results in a sintering of the very small clusters (if present) into larger ones which are detectable with TEM.

1.7.2 Inhomogeneous Distribution of Small Particles

When one deals with inhomogeneous distributions of small particles, such as catalytic metal particles in pellets (platinum in pellets of alumina, prepared by impregnation of the complete pellet resulting in a much higher platinum density near the

surface of the pellet) or an inhomogeneous size, shape, and composition distribution of precipitates (e.g., aluminum-containing Si–Mg precipitates), ultramicrotomy may be a good approach (see Sec. 1.8.1 of this Chapter).

1.7.3 Collection of Small Particles which are Finely Dispersed in a Fluid or Air

The collection of small particles finely dispersed in a fluid onto some sort of support can lead to artefacts, because of specimen support interaction and because of agglomeration. In this case one could consider the ultramicrotomy of fast-frozen specimens, as is done routinely in cryo-electron microscopy of biological specimens. As with biological specimens, using this method the 'natural' surroundings of the particles and their dispersion are maintained.

In the case of loose particles in air, these first have to be collected. Unlike dispersions in a fluid, this collection has to be done on some support or in a fluid, which is likely to result in a nonrepresentive sampling. Once the particles have been collected they can be embedded, for example in epoxy. An improved strength or density can be obtained by curing the epoxy at an elevated temperature either in a vacuum (to extract the air bubbles) or under pressure. For porous materials a complete impregnating procedure has to be applied. The epoxy 'ensemble' can be ultramicrotomed.

1.7.4 Various Special Applications

The area of particle analysis is so broad that clearly it is not possible to categorize all the special applications in this contribution. In the following, a few interesting examples (by no means comprehensive) of special applications are mentioned.

In a few cases when the particles are extremely homogeneous in all respects, these can settle into a three-dimensional superstructure [153]. Such superstructures may be interesting model systems to study hexagonal and cubic close-packed structures of large molecules. An example of the spontaneous close-packed stacking of particles is shown in Fig. 18.

In special cases where one is concerned with macromolecules (e.g., proteins) it is

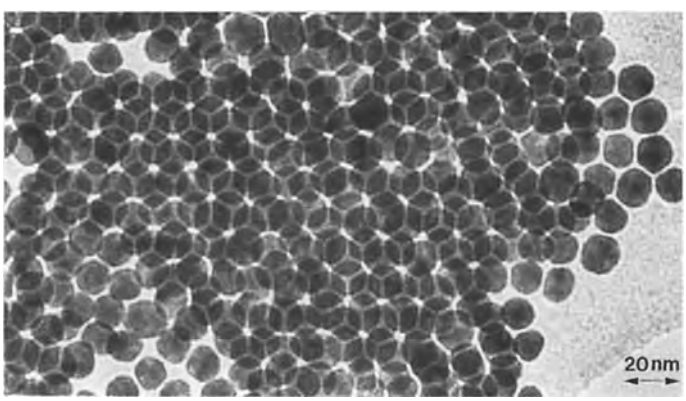


Figure 18. Example of the spontaneous stacking of large platinum clusters in an almost perfect close-packed arrangement on a carbon thin film, even though the clusters show a variation in size. This arrangement was obtained by dissolving a large agglomerate of these clusters which was already positioned on the carbon thin film. In the agglomerate the clusters did not show any ordering. The clusters were set free from the agglomate by dissolving the skin of large organic molecules (sodium sulfanifate).

possible to crystallize these into a two-dimensional crystal at a liquid–liquid interface [154]. The advantage of this technique is that it becomes possible to perform electron diffraction, thus extracting average structure information even from very beam-sensitive specimens.

It is frequently necessary to use image processing to extract useful information from the often noise-dominated electron micrograph. An ingenious example of image processing is reported in the literature [155], where 12- and 13-fold rotational symmetry is detected by rotational image averaging over a number of individual macromolecules.

In some cases one can circumvent the particle–substrate interaction by performing electron diffraction on a beam of supersonic clusters [156]. In these experiments it is possible to extract information on the mean size, the morphology, and the mean cluster temperature.

1.8 Specimen Preparation Related Topics

1.8.1 Preparation of Specimens and Supports

Four ways of sample preparation are employed: (1) crushing resulting in small, sometimes wedge-shaped fragments with a thin edge; (2) chemical polishing with a jet-stream of polishing liquid; (3) ion milling in which ions are focused onto a specimen under a low angle thus sputtering atoms and conglomerates of atoms from the surface; and (4) ultramicrotomy in which very thin slices are cut from the sample. The first technique is generally applied to some homogeneous bulk materials and can almost never be used for the study of grain boundaries and interfaces and for materials which change under the pressure of the crushing (metals). With this method the thin area is very small. The second and third techniques, presently most widely used in materials science, can provide very thin specimens. However, an inherent problem of these techniques is the presence of a layer (amorphous or crystalline) on the surface which is different from the 'real' material. The fourth method has as major disadvantages the creation of artefacts and the method of collection of the specimen, which is still done on a water surface and can result in changes at the surface of the slices. However, most artefacts created by ultramicrotomy can be identified as being due to this technique.

The surface contaminations can be divided into two groups: those due to the thinning method and those due to a reaction with the atmosphere. To minimize the first type of contamination one has to choose the best thinning method. To minimize the reaction with the atmosphere one has to perform the thinning and insertion of the specimen into the microscope in a controlled atmosphere (see Sec. 1.8.6).

Ultramicrotomy is widely used for making cuts of biological materials, and is optimized in particular for these types of specimens, but it is increasingly being used by materials scientists, in particular for polymers and catalysts [157]. It might be necessary to change the collection of the cut slices, which is routinely done on a water surface, since many specimens in materials science are sensitive to water. Cutting can be performed in a controlled atmosphere, as is done routinely

in the cryo-ultramicrotomy of biological specimens.

1.8.2 Passification of Small Metal Particles

Quite a number of catalysts consist of small metal particles, which are only stable in a reducing atmosphere. With most electron microscopes, however, one can only insert the specimen into the microscope by exposing the specimen to air. If this exposure is done instantaneously, most metal particles will be completely oxidized. In fact, such a catalyst can catch fire when exposed to air. This complete oxidation can be prevented by a process which is called passification: the catalyst is exposed to an atmosphere in which the oxygen partial pressure is slowly increased. A typical example of a too long exposure to air is shown in Fig. 19. This figure shows a nickel particle which is covered with an NiO film, due to reaction with air.

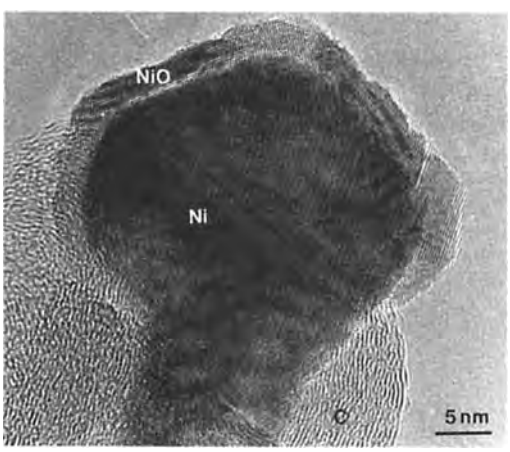


Figure 19. HREM image of a nickel particle on a carbon wire, showing the oxidation of the nickel surface due to exposure to air. Notice that no oxidation has taken place on the Ni–C interface.

1.8.3 Preparation of Holey Films and Very Thin (2 nm) Carbon Foils

Since the availability of a good thin foil support is essential to much catalyst research, we will briefly describe the preparation of thin holey films and very thin carbon films.

First, 0.2 g of Triafol is dissolved in 100 ml of ethyl acetate. After stirring for 1 h the nondissolved Triafol is separated by filtration. Next, approximately 50 ml of ethyl acetate is added to reduce the Triafol concentration (this is an essential step because if a fully saturated Triafol solution is used one may obtain a holey film with poorly developed holes). To this solution about 0.25 ml of glycerol is added, after which the solution is stirred vigorously for 30–60 min in combination with ultrasonic agitation. Next, carefully cleaned microscope glass slides are dipped into the solution. The film present on the microscope slide is floated off on a water–air interface by slowly lowering the slide under an angle of 20–45° into water. Grids are placed on the film. The whole ensemble of grids and film is removed from the water surface using filter paper with small pores. Finally, a thin carbon layer is deposited onto the ensemble, after which the grids can be taken off for storage.

For the preparation of a very thin carbon film, one starts with the holey film grids, described in the previous paragraph, but with a rather thick carbon layer deposited (the filter paper is rather gray in this case). The Triafol of the holey film grid is

dissolved, by dipping the grid in ethyl acetate. This grid is placed on a thin continuous Triafol film floating on a water surface (prepared as above but without the addition of glycerol). The grid–film ensemble is removed from the water surface by filter paper. After drying, a very thin carbon layer is deposited (in this case the filter paper has not visibly changed in color). Finally, the Triafol film is removed by ethyl acetate.

Evidently, with the present techniques of thin film deposition other types of films can be obtained by changing the carbon deposition step. This allows, for instance, the preparation of very thin films of germanium or MoSi which can be used for the determination of the information limit of the electron microscope. Another film which is more useful for catalysis research is the SiO_2 film.

1.8.4 Stability of Holey Films and Other Supports on (ex situ or in situ) Heating

If one is interested in performing TEM or HREM on exactly the same area with intermittent heating in specific atmospheres, a support is needed which is stable in these atmospheres and at these temperatures. We have experience with four types of supports:

(i) A holey carbon film on a copper or gold grid. If the temperature is higher than about 150°C a copper grid can no longer be used because copper from the grid diffuses (probably surface diffusion) into the carbon grid, thus forming a carbon grid covered with small copper or CuO particles. For this reason, it is much better to use a gold grid. The use of a gold grid is also recommended because gold is very inert in almost all catalytic reactions, such that it does not cause side-effects. One has to be aware of impurities in the gold grid, as mentioned by Datye et al. [158]. We have done reduction experiments in H_2 and H_2S up to 600°C without a noticeable change in the holey carbon film, except for very thin carbon films of about 2 nm.

(ii) A grid like that above but changed in such a way that a thin silicon layer is deposited under an N_2 atmosphere onto the carbon holey film grid, after which the grid is annealed at 400°C in air to burn off the carbon and the polymer [159]. Although the N_2 atmosphere was intended to make an Si_3N_4 film, such a film was not completely formed, which was evident from the still good electrical conductivity of the film (no charging occurred) and the rather low nitrogen signal in an EDX spectrum. Such films allow heating in air or oxygen up to 600°C.

(iii) A gold or platinum grid without any film. Particles are in this case supported by the grid itself. With this method the only limitation is that the number of particles in positions which allow TEM is small, in particular if one also wants to perform EDX element analysis, since about half of the particles are shielded from the detector by the grid.

(iv) A thin Si_3N_4 membrane on a silicon wafer with a window in the silicon wafer [160]. These membranes are

also stable against heating. A disadvantage is that the silicon wafer is rather thick (about 1 mm), making it less usable for microscopes with a small gap between the objective pole pieces.

For catalysis research several other substrates have also been reported: thin substrates have been made by ion milling or dimpling [161]; and thin planar films of an oxide are formed on a substrate, after which the substrate is removed [162], similar to the preparation of holey carbon films as discussed in Section 1.8.3 of this Chapter.

The advantage of the first four methods is that real catalytic materials can be used, whereas for the latter methods one uses the thin film or thinned area as the substrate. Another disadvantage is that the geometry of the particles on the substrate along the viewing direction does not allow imaging of the particle–substrate interface.

1.8.5 In situ Transmission Electron Microscopy

If one wants to study the surface in the state in which it was applied, one needs to create this environment in the characterization equipment. This is relatively easy to obtain in scanning tip microscopes but is generally impossible for electron microscopes. The strong interaction of electrons with atoms requires a low gas pressure. In particular, in the case of catalysts it is almost impossible to approach the real conditions, because the catalytic reactions are performed at high pressures. One can try to approach the required environment in several ways: (i) by placing an environmental cell inside the microscope or (ii) by performing all stages of specimen preparation and specimen transfer in an atmosphere under which the surface of the catalyst is not significantly changed from the arrangement during the catalytic reaction.

1.8.5.1 An Environmental Cell

Several transmission electron microscopes allow the mounting of an environmental cell between the pole pieces of the objective lens. Since this requires a large separation of the pole pieces such a cell does not allow HREM. Also, the gas inside the cell does not allow a high resolution. The gas pressure is still quite far from the conditions under which the catalysts were applied. Still, interesting studies have been done [163], in particular the sintering of metal particles and the effect of cycles of oxidation and subsequent reduction on the shape and size of metal–metal oxide particles [164]. Parkinson [165] obtained a resolution of 0.3 nm with an environmental cell with a pressure of 2.6×10^6 Pa of nitrogen. Doole et al. [166] reported an in situ cell with pressures up to 2×10^4 Pa and temperatures up to 700°C with the maintenance of good resolution.

1.8.5.2 Specimen Handling and Specimen Transfer in a Controlled Atmosphere

An alternative to an environmental cell is a method in which the catalytic reaction is performed outside the transmission electron microscope in a dedicated reactor and

with a transfer system which allows transport of the specimen from the reactor into the microscope and vice versa. This requires a transfer system allowing control of the atmosphere such that it does not cause a significant change to the catalyst particles. Obviously, one limitation of this method is the vacuum of the microscope; thus the transfer system does not require a considerably better vacuum than that of the microscope. The main advantage of this method is that the reactions can be performed under real reaction conditions. The transfer system is discussed in Sec. 1.8.6 of this Chapter.

1.8.6 Specimen Transfer under Protective Atmospheres

Although most electron microscopy techniques have become well established and appreciated, the insertion of specimens is still very primitive. Except for a few TEM and dedicated STEM instruments the transfer of specimens is done through air. The transfer through air can create serious artefacts, which will result in a situation completely different from that which one would like to study. With the high-resolution electron microscopes used nowadays even a monoatomic layer can be seen; thus, when one aims to study surface-related processes the surface must be as close as possible to its state in the catalytic reaction. Also, since for HREM very thin specimens (less than 10 nm is often the best experimental condition) are needed, even a thin surface layer of 1–2 nm (which is 10–30% of the thin foil in the viewing direction) can have a devastating effect on the images.

In the study of catalytic metal particles, the metal particles have to be passified before transport to the transmission electron microscope, by slowly increasing the partial pressure of oxygen, resulting in a thin oxide layer thus preventing further oxidation. This will produce a quite different state to that of the catalytic reaction. If this is not done carefully or if the time of exposure to air is too long, a rather thick reaction layer can be observed, as is shown in Fig. 19.

In Delft we have developed specimen-handling systems which allow transfer in a controlled atmosphere, and a specimen manipulation chamber which allows the preparation and transfer of specimens under controlled conditions; of particular interest are several specimen holders, for example the double-tilt holder, which can be used in an electron microscope with an objective pole piece separation of about 6 mm or more (e.g., Philips CM30T) and the single-tilt holder, requiring an objective pole piece separation of only 2.3 nm (Fig. 20). Other less restrictive precautions

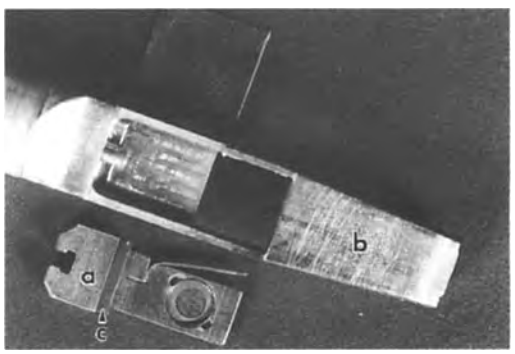


Figure 20. One of the vacuum sample holders developed in Delft. The specimen is positioned in a 'drawer' (a), which can be pushed into the tip (b) of the holder and can be closed air-tight by a flat piece of Viton (c). The mechanical stability of this holder is as good as that of a standard single-tilt holder.

are sometimes sufficient. For instance, Zhang et al. [167] made a box around the microscope with an over-pressure of nitrogen to prevent air leaking in.

1.8.7 Surface Clean Specimens

Chemical polishing with a jet-stream of polishing liquid, and ion milling in which ions are focused on a specimen under a low angle thus shooting atoms and conglomerates of atoms from the surface, presently most widely used in materials science, can provide very thin specimens. However, a problem inherent to these techniques is the occurrence of a layer (amorphous or crystalline) on the surface which is different from the 'real' material. The formation of a contamination layer can occur because of damage to the lattice by the thinning medium, redeposition on the surface of material removed from another area, and the deposition of reaction products. Lattice damage will occur predominantly with ion milling. This can be minimized by a final ion milling step at low ion energy (2 keV) and with a very small incidence angle. Redeposition forming an amorphous surface layer can also occur with ion milling. In ion milling, an amorphous layer is partly removed by finishing the process with a low ion energy and incidence angle. The deposition of reaction products will occur mainly with electropolishing. This can be reduced by using an optimized polishing liquid and minimizing the time interval between the actual polishing and the removal of the polishing liquid after the polishing is stopped. However, although the presence of an amorphous layer can be minimized by careful working, it cannot be prevented, especially when the composition or the structure of the specimen is complex (e.g., $Bi_2Sr_2CaCu_2O_8$ or grain boundaries of more simple structures).

An additional problem of TEM investigations of thinned specimen is the contamination which occurs in the electron beam. This contamination is in particular a problem when small probe sizes are used. Thus, although field emission guns as the electron source make it feasible to work with probe sizes in the range 1–10 Å, allowing the analysis of very small defects and inclusions, this analysis is often hampered by a fast contamination. With such small probe sizes the contamination can grow rapidly, thus obscuring the HREM image very quickly (within a few seconds), making it impossible to check whether or not the electron beam is still focused at the same position on the sample.

To create as little damage as possible in the top layer of the etched specimen it is necessary to transfer the minimum amount of energy to the surface atoms of the specimen. Theoretically this low-energy transfer can be achieved by directing a high energy (kiloelectronvolt) ion beam at glancing angles on the specimen. In this geometry most of the ion energy is carried away by the scattered ions. However, because of the presence of steps on the surface of the specimen a considerable fraction of the ions will enter the sample and may penetrate deeply. Low-energy transfer can also be achieved by using low-energy ions with perpendicular incidence. By doing so the problem of penetration of fast particles into the sample can be avoided.

Theoretical models for sputtering and damage production by impinging particles

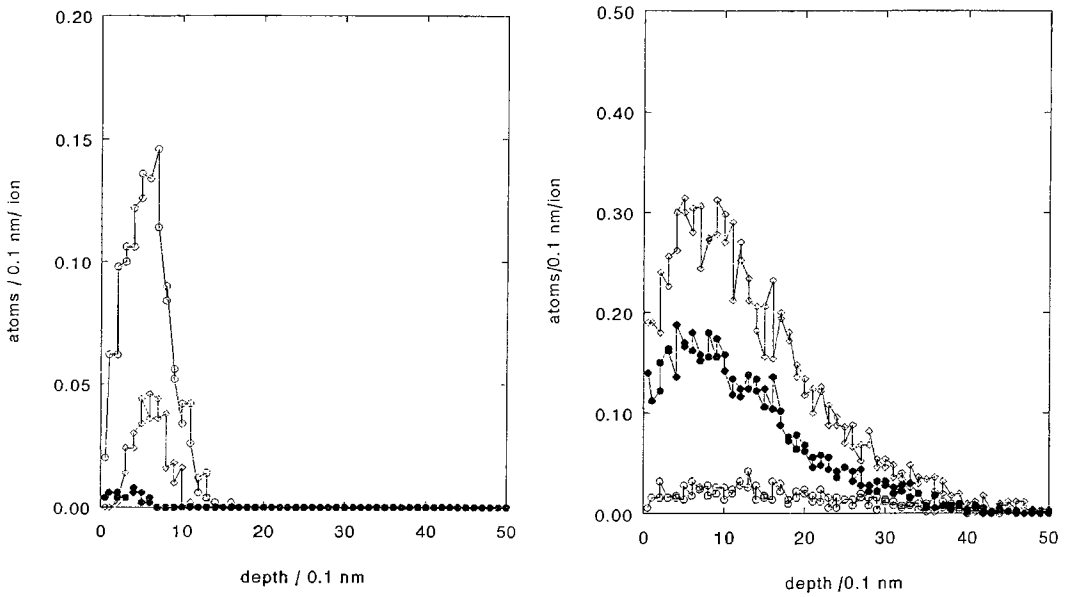


Figure 21. Range distributions of implanted and displaced atoms in $YBa_2Cu_3O_7$ calculated using the TRIM software. Depth profiles are shown of neon, copper, and oxygen (open circles, open squares, and full circles, respectively) when the high-temperature superconductor $YBa_2Cu_3O_7$ is irradiated with 40 eV perpendicularly incident Ne^+ ions (left) and 1 keV ions at an angle of 20° to surface (right).

(using the TRIM software package) indicate that it is possible to minimize the defect layer to less than 0.5 nm if light ions with an energy of about 40 eV are used. Low-energy plasma thinning and plasma cleaning are used in the fabrication of electronic devices. In order to create as little damage as possible to the remaining material, plasma energies of 50–200 eV are mostly applied. Examples of calculations for $YBa_2Cu_3O_7$ are shown in Fig. 21. There is a dramatic reduction in the range and number of displaced copper atoms when 40 eV rather than 1 keV ions are used. This occurs to the same extent for other cations. These calculations show that a low-energy sputtering treatment will preserve the structure much better. At 1 keV an amorphous layer of at least 3 nm is expected.

1.9 References

[1] J. A. Moulijn, *Catalysis, An Integrated Approach to Homogeneous Heterogeneous and Industrial Catalysis*, Elsevier, Amsterdam **1993**.

[2] R. W. Siegel, *Mater. Sci. Eng.* **1993,** *A168*, 189; T. Gleiter, *Progr. Mater. Sci.* **199X,** *33*, 223.

[3] H. Haberland (Ed.), *Clusters of Atoms and Molecules*, Springer-Verlag, Berlin **1994**; K. J. Klabunde, *Free Atoms, Clusters and Nanoscale Particles*, Academic Press, San Diego **1994**.

[4] L. M. Falikov, D. T. Pierce, S. D. Bader, R. Gronsky, K. B. Hathaway, H. J. Hopster, D. N. Lambeth, S. S. P. Parkin, G. Prinz, M. Salamon, I. K. Schuller, R. H. Victoria, *J. Mater. Res.* **1990,** *5*, 1299.

[5] R. M. Stockmann, H. W. Zandbergen, A. D. van Langeveld, J. A. Moulijn, *J. Molec. Catal. A* **1995**, *102*, 147; S. Eijsbouts, J. J. L. Heinerman, H. J. W. Elzerman, *Appl. Catal.* **1993,** *A105*, 53.

[6] G. W. Nieman, J. R. Weertman, R. W. Siegel, *J. Mater. Res.* **1991,** *6*, 1012.

[7] V. R. Choudhari, S. K. Choudhari, A. N. Gokarn, *Ind. Eng. Chem. Res.* **1989,** *28*, 33.

[8] I. Dutta, S. M. Allen, *J. Mater. Sci. Lett.* **1991,** *10*, 323.
[9] G. N. van den Hoven, E. Snoeks, A. Polman, J. W. M. van Uffelen, Y. S. Oei, M. K. Smit, *Appl. Phys. Lett.* **1993,** *62*, 3065.
[10] S. S. P. Parkin, N. More, K. P. Roche, *Phys. Rev. Lett.* **1990,** *64*, 2304.
[11] M. F. H. van Tol, M. A. Quinlan, F. Luck, G. A. Somorjai, B. E. Nieuwenhuys, *J. Catal.* **1991,** *129*, 186.
[12] H. Hahn, J. Logas, T. Bier, R. A. Averback, *Mater. Res. Soc. Symp. Proc.* **1989,** *132*, 32.
[13] R. R. Bhave (Ed.), *Inorganic Membranes Synthesis, Characterisation and Application*, van Nostrand Reinhold, New York **1991**.
[14] M. Murakami, *NATO ASI Ser. E* **1994,** *263*, 433.
[15] R. Ruther et al., *Thin Solid Films* **1994,** *251*, 30.
[16] B. G. Demczyk, *Ultramicroscopy* **1992,** *47*, 425; C. Huang, D. E. Laughin, P. O. Mitchell, J. E. Snyder, J. O. Artman, *J. Magn. Magn. Mater.* **1986,** *54–57*, 1676.
[17] V. Ponec, *Adv. Catal.* **1983,** *32*, 149.
[18] V. Ponec, *NATO ASI Ser. B* **1988,** *191*, 301.
[19] H. Weller, A. Fojtik, A. Henglein, *Chem. Phys. Lett.* **1985,** *117*, 485.
[20] S. S. P. Parkin, N. More, K. P. Roche, *Phys. Rev. Lett.* **1990,** *64*, 2304.
[21] W. B. Innes in *Catalysis* (Ed.: P. H. Emmett), Reinhold, New York **1995,** p. 1.
[22] T. Koets, W. Welters, R. A. van Santen, L. E. Y. Nonneman, V. Ponec, *Studies Surf. Sci.* **1990,** *61*, 238; B. J. Kip, P. A. T. Smeets, J. van Wolput, H. W. Zandbergen, J. van Grondelle, R. Prins, *Appl. Catal.* **1987,** *33*, 157.
[23] K. J. Williams, M. E. Levin, M. Salmeron, A. T. Bell, G. A. Somorjai, *Catal. Lett.* **1988,** *2*, 331.
[24] V. R. Choudhari, S. K. Choudhari, A. N. Gokarn, *Ind. Eng. Chem. Res.* **1989,** *28*, 33.
[25] S. D. Mikhailenko, A. B. Fasman, N. A. Maksimova, E. V. Leongard, E. V. Shapiro, G. V. Antoshin, *Appl. Catal.* **1984,** *12*, 141.
[26] G. W. Nieman, J. R. Weertman, R. W. Siegel, *J. Mater. Res.* **1991,** *6*, 1012.
[27] R. R. Bhave (Ed.), *Inorganic Membranes Synthesis, Characterisation and Application*, van Nostrand Reinhold, New York **1991**.
[28] R. R. Bhave, *J. Catal.* **1990,** *126*, 324.
[29] N. S. Stoloff, V. K. Sikka (Eds)., *Physical Metallurgy and Processing of Intermetallic Compounds*, Chapman and Hall, New York **1996**.
[30] P. P. J. Maziasz, *J. Met.* **1989,** *41*, 14.
[31] D. P. Pope, S. S. Ezz, *Int. Met. Rev.* **1984,** *29*, 136.
[32] P. Hirsch, M. Kässens, L. Reimer, R. Senkel, M. Spranck, *Ultramicroscopy* **1993,** *50*, 263.
[33] D. C. Joy, D. E. Newbury, D. L. Davidson, *J. Appl. Phys.* **1982,** *53*, R81; J. A. Venables, C. J. Harland, *Phil. Mag.* **1973,** *27*, 1193; N. C. Krieger Lassen, D. Juul Jensen, K. Conradsen, *Scanning Microsc.* **1992,** *6*, 115.
[34] D. A. Porter, K. E. Easterling, *Phase Transformations in Metals and Alloys*, 2nd edn., Chapman and Hall, New York **1992/1993**.
[35] F. Delannay (Ed.), *Characterization of Heterogeneous Catalysts*, Marcel Dekker, New York **1983**.
[36] L. A. Freeman, A. Howie, M. M. J. Treacy, *J. Microsc.* **1977,** *111*, 165.
[37] W. Krakow, L. A. Howland, *Ultramicroscopy* **1976,** *2*, 33.
[38] A. J. Bleeker, P. Kruit, *Rev. Sci. Instrum.* **1991,** *62*, 350.
[39] P. Kruit, J. A. Venables, *Ultramicroscopy* **1988,** *25*, 183.
[40] S. J. Pennycook, D. E. Jesson, *Ultramicroscopy* **1991,** *37*, 14.
[41] L. Reimer, *Springer Ser. Opt. Sci.* **1993,** *36*, 122, 202, 221.
[42] A. Howie, *NATO ASI Ser. B* **1988,** *191*, 11, 19.
[43] M. M. J. Treacy, S. B. Rice, *J. Microsc.* **1989,** *156*, 211; S. J. Pennycook, D. E. Jesson, *Ultramicroscopy* **1991,** *37*, 14.
[44] S. J. Pennycook, *Ultramicroscopy* **1991,** *37*, 256.
[45] Z. Ma, C. Chu, J. Shen, S. Pang, *Ultramicroscopy* **1992,** *42–44*, 1350.
[46] S. S. Sheiko, M. Möller, E. M. C. M. Reuvekamp, H. W. Zandbergen, *Phys. Rev. B.* **1993,** *48*, 5675; S. S. Sheiko, M. Möller, E. M. C. M. Reuvekamp, H. W. Zandbergen, *Ultramicroscopy* **1994,** *53*, 371.
[47] N. D. Lang in *Scanning Tunnelling Microscopy III* (Eds.: R. Wiesendanger, H. J. Güntherodt), Springer-Verlag, Hamburg **1993**.
[48] Wouda, B. Nieuwenhuis, *Ultramicroscopy* **1993,** *42*, 110.
[49] J. Winttlerlin, R. J. Behm in, *Scanning Tunnelling Microscopy I* (Eds.: R. Wiesendanger and H. J. Güntherodt), Springer-Verlag, Hamburg **1991**.
[50] G. Ertl, J. Küppers, *Low Energy Electrons and Surface Chemistry*, VCH, Weinheim **1985**; B. Feuerbacher, B. Filton, R. F. Willis (Eds.), *Photoemission and Electronic Properties of Surfaces*, Wiley, New York **1978**.
[51] D. Briggs, M. P. Seah (Eds.), *Practical Surface Analysis by Auger and X-Ray Photoemission Spectroscopy*, Wiley, New York **1983**.
[52] J. C. Vickerman, A. Brown, N. M. Reed (Eds.), *Secondary Ion Mass Spectroscopy, Principles and Applications*, Clarendon Press, Oxford **1989**.

[53] A. Benninghoven, F. C. Rüdenauer, H. W. Werner, *Secondary Ion Mass Spectroscopy, Basic Concepts, Instrumental Aspects, Applications and Trends*, Wiley, New York **1987**.
[54] J. L. Falconer, J. A. Schwartz, *Cat. Rev.-Sci. Eng.* **1983,** *25*, 141.
[55] T. E. Cranshaw, B. W. Dale, G. O. Longworth, C. E. Johnson, *Mössbauer Spectroscopy and its Applications*, Cambridge University Press, Cambridge **1985**.
[56] A. M. van der Kraan, J. W. Niemantsverdriet in *Industrial Applications of the Mössbauer Effect* (Eds.: G. J. Long, J. G. Stevens), Plenum Press, New York **1985,** p. 609.
[57] Y. J. Yiu, *Infrared Spectroscopy* **1989,** 146.
[58] A. Bianconi, L. Incoccia, S. Stipcich (Eds.), *EXAFS and Near Edge Structure*, Springer-Verlag, Berlin **1983**.
[59] R. Birringer, U. Herr, H. Gleiter, *Suppl. Trans Jpn. Inst. Met.* **1986,** *27*, 43.
[60] G. Porod, in *Small Angle X-Ray Scattering* (Eds.: O. Glatter, O. Kratky), Academic Press, London **1982,** p. 17.
[61] M. R. Fitzsimmons, S. L. Sass, *J. Phys.* **1988,** *49*, C5.
[62] A. T. Bell, A. Pines (Eds.), *NMR Techniques in Catalysis*, Marcel Dekker, New York **1994**; J. W. Hennel, J. Klinowski, *Fundamentals of Nuclear Magnetic Resonance*, Longman, Harlow **1993**.
[63] G. Schmid, *Polyhedron* **1988,** *7*, 2321.
[64] O. Terasaki, *J. Electr. Microsc.* **1994,** *43*, 337.
[65] M. E. Davis, R. F. Lobo, *Chem. Mater.* **1992,** *4*, 756.
[66] C. T. Kresge, M. E. Leonowitz, W. J. Roth, J. C. Vartuli, J. S. Beck, *Nature* **1992,** *359*, 710; J. S. Beck, J. C. Vartuli, W. J. Roth, M. E. Leonowitz, C. T. Kresge, K. D. Schmitt, C. T. Chu, D. H. Olson, E. W. Sheppard, S. B. McCullen, J. B. Higgens, J. L. Schlenker, *J. Am. Chem. Soc.* **1992,** *114*, 10834.
[67] A. Fitouri, D. Kumar, L. M. Bull, T. Besier, P. Sieger, Q. Huo, S. A. Walker, J. A. Zasadzinski, C. Glinka, J. Nicol, D. Margolese, G. D. Stcky, B. F. Chmelka, *Science* **1995,** *267*, 1138.
[68] R. M. Stockmann, H. W. Zandbergen, A. D. van Langeveld, R. Prins, J. A. Moulijn, unpublished results.
[69] A. Thölén, *Mater. Sci. Eng. A* **1993,** *168*, 131.
[70] R. W. Siegel, *Mater. Sci. Eng.* **1993,** *A168*, 189.
[71] H. Gleiter, *Proc. Mater. Sci.* **1989,** *33*, 223; M. Nastasi, D. M. Parker, H. Gleiter (Eds.), *Mechanical Properties and Deformation Behaviour of Materials Having Ultra-fine Microstructures*, Kluwer, Dordrecht **1993**.
[72] G. W. Nieman, J. R. Weertman, R. W. Siegel, *J. Mater. Res.* **1991,** *6*, 1012.
[73] V. R. Choudhari, S. K. Choudhari, A. N. Gokarn, *Ind. Eng. Chem. Res.* **1989,** *28*, 33.
[74] Z. Li, S. Ramasamy, H. Hahn, R. W. Siegel, *Mater. Lett.* **1988,** *6*, 195.
[75] H. A. Atwater, K. V. Shehelov, S. S. Wong, K. J. Vahala, R. C. Flagan, M. L. Brongersma, A. Polman, *Mater. Res. Soc. Proc.* **1994,** *316*, 409.
[76] I. Dutta, S. M. Allen, *J. Mater. Sci. Lett.* **1991,** *10*, 323.
[77] R. Benedictus, K. Han, A. Boettger, H. W. Zandbergen, E. J. Mittemeijer in *Solid–Solid Phase Transformations* (Eds.: Johnson et al.) **1994,** p. 1027.
[78] Z. Li, *Progr. Mater. Sci.* **1989,** *33*, 223.
[79] L. L. Hench, D. R. Ulrich, *Science of Chemical Processing*, Wiley, New York, **1986**; Uyeda, *Prog. Mater. Sci.* **1991,** *35*, 1.
[80] K. Ganapathy, D. A. Ridley, *Scripta Metall. Mater.* **1990,** *24*, 1675; Bakker et al.
[81] G. N. van den Hoven, E. Snoeks, A. Polman, J. W. M. van Uffelen, Y. S. Oei, M. K. Smit, *Appl. Phys. Lett.* **1993,** *62*, 3065.
[82] U. Herr, J. Jing, U. Gonser, H. Gleiter, *Solid State Commun.* **1992,** 463.
[83] H. Gleiter, *Prog. Mater. Sci.* **1989,** *33*, 223.
[84] M. R. Fitzsimmons, S. L. Sass, *J. Phys. C.* **1988,** *49*, 5.
[85] J. D. Verhoeven, *Fundamentals of Physical Metallurgy*, Wiley, New York **1975,** p. 363.
[86] J. D. Verhoeven **1975**; D. A. Porter, K. E. Easterling, *Phase Transformations in Metals and Alloys*, 2nd edn, Chapman and Hall, New York **1992/1993**, pp. 149, 263.
[87] M. Murakami, *NATO ASI Ser. E* **1994,** *263*, 433.
[88] U. Messerschmidt, M. Batsch, *Ultramicroscopy* **1994,** *56*, 225; E. P. Bullen, K. F. Hale, *Practical Methods in Electron Microscopy*, North-Holland, Amsterdam **1981,** Vol. 9.
[89] L. M. Falikov, D. T. Pierce, S. D. Bader, R. Gronsky, K. B. Hathaway, H. J. Hopster, D. N. Lambeth, S. S. P. Parkin, G. Prinz, M. Salamon, I. K. Schuller, R. H. Victoria, *J. Mater. Res.* **1990,** *5*, 1299.
[90] J. L. Dormann, D. Fiorani (Eds.), *Proc. Int. Workshop on Studies of Magnetic Properties of Fine Particles and their Relevance to Materials Science, Rome, Italy, November 4–8, 1991*, Elsevier, Amsterdam **1992**.
[91] R. H. Victoria, *Ultramicroscopy* **1992,** *47*, 318.
[92] M. N. Baibich, J. M. Broto, A. Fert, F. Nguren Van Dau, F. Pettroff, P. Eilenne, G. Creuzet, A. Friederich, J. Chazelas, *Phys. Rev. Lett.* **1988,** *61*, 2472.
[93] T. Katayaman, H. Awano, Y. Nishihara, *J. Phys. Soc. Jpn.* **1986,** *44*, 2539.

[94] P. F. Carcia, *J. Appl. Phys.* **1988,** *63*, 5073.
[95] Z. G. Li, P. F. Garcia, *Ultramicroscopy* **1992,** *47*, 313.
[96] M. Hartmann, B. A. J. Jacobs, J. J. M. Braat, *Philips Tech. Rev.* **1985,** *42*, 37.
[97] S. S. P. Parkin, N. More, K. P. Roche, *Phys. Rev. Lett.* **1990,** *64*, 2304.
[98] Y. Martin, H. K. Wickramasignhe, *Appl. Phys. Lett.* **1987,** *50*, 1455; J. J. Seanz, N. Garcia, P. Grütter, E. Meyer, H. Heinzelmann, R. Wiesendanger, L. Rosenthaler, H. R. Hidber, H. J. Güntherrodt, *J. Appl. Phys.* **1987,** *62*, 4293.
[99] P. Grütter, H. J. Mamin, D. Rugar in *Scanning Tunneling Microscopy* (Eds.: H. J. Güntherodt, R. Wiesendanger), Springer-Verlag, Berlin **1992.**
[100] G. Persch, H. Strecker, *Ultramicroscopy* **1992,** *42–44*, 1269; P. Grütter, D. Rugar, H. J. Mamin, *Ultramicroscopy* **1992,** *47*, 393.
[101] S. D. Bader, *J. Magn. Magn. Mater.* **1991,** *100*, 440.
[102] M. R. Scheinfein, P. J. Ryan, J. Ungaris, D. T. Pierce, R. J. Celotta, *Appl. Phys. Lett.* **1990,** *57*, 1817.
[103] A. Tonomura, *Ultramicroscopy* **1992,** *47*, 419.
[104] P. L. Gammel, D. J. Bishop, G. J. Dolan, J. R. Kwo, C. A. Murray, L. F. Schneemeyer, J. V. Waszczak, *Phys. Rev. Lett.* **1987,** *59*, 2592.
[105] A. S. Ramachandran, S. L. Anderson, A. K. Datye, *Ultramicroscopy* **1993,** *51*, 282.
[106] D. G. Jensen, *J. Phys. D: Appl. Phys.* **1995,** *28*, 549.
[107] J. M. Zuo, J. C. H. Spence, *Ultramicroscopy* **1991,** *35*, 185.
[108] K. Tsuda, M. Tanaka, *Acta Crystallogr. A* **1995,** *7*, 000.
[109] B. D. Sha, H. F. Fan, F. H. Li, *Acta Crystallogr. A* **1993,** *49*, 877.
[110] J. Jansen, D. Tang, H. Schenk, H. W. Zandbergen, unpublished results.
[111] J. Jansen, D. Tang, H. W. Zandbergen, H. Schenk, *Acta Crystallogr.* (submitted).
[112] F. Delannay, *Catal. Rev. Sci. Eng.* **1980,** *22*, 141.
[113] V. D. Scott, G. Love (Eds.), *Quantitative Electron-Probe Microanalysis*, Ellis Horwood/John Wiley, Chichester **1983.**
[114] F. Hofer, P. Warbicler, W. Grogger, *Ultramicroscopy* **1995,** *59*(1–4), 15.
[115] G. Cliff, G. W. Lorimer, *J. Microsc.* **1975,** *103*, 203.
[116] C. E. Lyman, R. E. Lakis, H. G. Stengr, *Ultramicroscopy* **1995,** *58*(1), 25.
[117] L. D. Marks, *Surf. Sci.* **1984,** *139*, 281.
[118] D. Tang, J. Jansen, H. W. Zandbergen, M. op de Beeck, D. van Dyck, *Ultramicroscopy* (in press).
[119] Jahanmir, B. G. Haggar, J. B. Hayes, *Scanning Microsc.* **1992,** *6*, 625.
[120] J. Wintterlin, R. J. Behm in *Scanning Tunnelling Microscopy III* (Eds.: R. Wiesendanger, H. J. Güntherodt), Springer-Verlag, Hamburg **1993.**
[121] D. F. Ogletre, C. Ocal, B. Marccon, G. A. Somojai, M. Salmeron, T. Beeke, W. Siekhaus, *J. Vac. Sci. Technol.* **1990,** *A8*, 297.
[122] G. F. A. van de Walle, H. van Kempen, P. Wyder, C. G. Flipse, *Surf. Sci.* **1987,** *181*, 27.
[123] H. Brune, J. Wintterlin, G. Erlt, R. J. Behm, *Europhys. Lett.* **1990,** *13*, 123; P. H. Lippel, R. J. Wilson, M. D. Miller, C. Wöll, S. Chiang, *Phys. Rev. Lett.* **1989,** *62*, 171.
[124] C. Gunther, R. J. Behm, *J. Appl. Phys.* **1987,** *62*, 3859.
[125] J. Wintterlin, R. Schuster, D. J. Coulman, G. Ertl, R. J. Behm, *J. Vac. Sci. Technol.* **1991,** *B9*, 902.
[126] J. L. Hutchison, N. Briscoe, *Ultramicroscopy* **1985,** *18*, 435.
[127] D. J. Smith, *NATO ASI Ser. B* **1988,** *191*, 43.
[128] J. A. Venables, C. J. Harland, *Phil. Mag.* **1973,** *27*, 1193.
[129] D. C. Joy, D. E. Newbury, D. L. Davidson, *J. Appl. Phys.* **1982,** *53*, R81; H. Nedrieg, *J. Appl. Phys.* **1982,** *53*, R15.
[130] Q. Liu, *Ultramicroscopy* **1995,** *60*(1), 81; A. Gemperle, J. Gemperlová, *Ultramicroscopy* **1995,** *60*, 207.
[131] L. R. Wallenberg, Ph.D. thesis, Lund, Sweden **1987.**
[132] S. Giorgio, C. Chapon, C. R. Henry, G. Nihoul, *Phil. Mag. B* **1993,** *67*(6), 773.
[133] B. Viguier, K. J. Hemker, G. Vanderschaeve, *Phil. Mag. A* **1994,** *69*(1), 19.
[134] A. Thölén, *Micron Microsc. Acta* **1991,** *22*(1/2), 187.
[135] A. Thölén, *Mater. Sci. Eng. A* **1993,** *168*, 131.
[136] M. J. Yacamàn, L. D. Romeau, A. Gómez, Z. A. Munir, *Phil. Mag. A* **1979,** *40*(5), 645.
[137] Q. Liu, *MRS Bull.* **1996,** *21*(2).
[138] J. Jahanmir, B. G. Haggar, J. B. Hayes, *Scanning Microsc.* **1992,** *6*, 625.
[139] M. Pan, P. A. Crozier, *Ultramicroscopy* **1993,** *48*, 332.
[140] Zandbergen, H.W., D. van Dyck in *Zeolites: Facts, Figures, Future* (Eds.: P. A. Jacobs, R. A. van Santen), Elsevier, Amsterdam **1989,** p. 599.
[141] C. J. Humphreys, T. J. Bullogh, R. W. Devenish, D. M. Maher, P. E. Turner, *Scanning Microsc. Suppl.* **1990,** *4*, 185.
[142] C. J. Humphreys, *Micron Microsc. Acta* **1991,** *22*(1/2), 147.
[143] L. R. Wallenberg, J.-O. Bovin, G. Schmid, *Surf. Sci.* **1985,** *156*, 256; L. R. Wallenberg, J.-O. Bovin, D. J. Smith, *Naturwissenschaften* **1985,** *72*, 539; L. D. Marks, *Rep. Prog. Phys.* **1994,** *57*, 603.

[144] P. R. Williams, *Phys. Rev. Lett.* **1987,** *61*, 1867.
[145] L. W. Hobbs, *Scanning Microsc. Suppl.* **1990,** *4*, 171.
[146] L. Reimer, *Springer Ser. Opt. Sci.* **1993,** *36*, 00.
[147] L. D. Marks, *Langmuir* **1994,** *10*, 368.
[148] L. D. Marks, *Surf Sci.* **1994,** 487.
[149] O. H. Nielsen, J. P. Sethna, P. Stoltze, K. W. Jacobsen, J. K. Nørskov, *Europhys. Lett.* **1994,** *26*, 51.
[150] J. Cazaux, *Ultramicroscopy* **1995,** *60*(3), 411.
[151] H. W. Zandbergen, C. W. R. van Engelen, J. H. C. van Hooff, *Appl. Catal.* **1986,** *25*, 231.
[152] D. J. Smith, *NATO ASI Ser. B* **1988,** *191*, 43.
[153] M. D. Bentzon, J. van Wonterghem, S. Mørup, A. Thölén, C. J. Koch, *Phil. Mag. B* **1989,** *60*(2), 169; X. F. Zhang, X. B. Zhang, B. Bernearts, G. van Tendeloo, S. Amelinckx, J. van Landuyt, H. Werner, *Ultramicroscopy* **1994,** *55*, 25.
[154] K. Aoyama, K. Ogawa, Y. Kimura, Y. Fujiyoshi, *Ultramicroscopy* **1995,** *57*(4), 354.
[155] E. Kocsis, M. E. Cerritelli, B. L. Trus, N. Cheng, A. C. Steven, *Ultramicroscopy* **1995,** *60*(2), 219.
[156] H. Haberland (Ed.), *Clusters of Atoms and Molecules*, Springer-Verlag, Berlin **1994**; K. J. Klabunde, *Free Atoms, Clusters and Nanoscale Particles*, Academic Press, San Diego **1994,** p. 231.
[157] T. F. Malis, D. Steele, *Mater. Res. Symp. Proc.* **1990,** *199*, 1.
[158] K. Datye, D. J. Smith, *Catal. Rev.* **1992,** *34*, 129.
[159] H. W. Zandbergen, A. Pruymboom, G. van Tendeloo, *Ultramicroscopy* **1988,** *24*, 45.
[160] J. W. M. Jacobs, D. Schrijvers, *J. Catal.* **1987,** *103*, 436.
[161] F. Rumpf, H. Poppa, M. Boudart, *Langmuir* **1988,** *4*, 722.
[162] E. Ruckenstein, Y. F. Chu, *J. Catal.* **1979,** *59*, 109; H. Glassl, K. Kramer, K. Hayek, *J. Catal.* **1981,** *68*, 388; H. Poppa, C. A. Papagoergopoulos, E. Bauer, *Ultramicroscopy* **1986,** *20*, 145.
[163] H. Mori, M. Kamatsu, H. Fujita, *Ultramicroscopy* **1993,** *51*, 31; R. Sinclair, T. J. Konno, *Ultramicroscopy* **1994,** *56*, 225.
[164] L. D. Marks, *Surf Sci.* **1984,** *139*, 281.
[165] G. M. Parkinson, *Catal. Lett.* **1989,** *2*, 303.
[166] R. C. Doole, G. M. Parkinson, J. L. Hutchison, M. J. Goringe, P. J. F. Harris, *Jeol News* **1992,** *30*, 30.
[167] X. F. Zhang, X. B. Zhang, B. Bernearts, G. van Tendeloo, S. Amelinckx, J. van Landuyt, H. Werner, *Ultramicroscopy* **1994,** *55*, 25.

Handbook of Microscopy
S. Amelinckx, D. van Dyck,J. van Landuyt , G. van Tendeloo

2 Structural Phase Transformations

2.1 Introduction

Ever since sorbite was named after Henry Clifton Sorby (1826–1908) and martensite after Adolf Martens (1850–1914) for their pioneering work in microstructural research on transformations in steels, analysis by microscopy has played a dominant role in the study of structural phase transformations. Because of the continuing importance of microscopy for the characterization of steels it is not surprising that one of the most comprehensive, systematic monographs treating phase transformations on the basis of microscopic evidence is entitled 'De Ferri Metallographia' [1]. In the present context it will not be possible to cover the subject systematically. Therefore a short overview of the main aspects of structural phase transformations and their classification will be followed by a number of examples.

2.2 Phase Transitions and Structural Phase Transformations

Phase transitions in the solid state can be associated with changes of several variables of state such as composition, crystal structure, magnetic and electrical polarization, and conductivity. The experimental study of phase transitions by structure and property measurements and their theoretical treatment are large fields of solid state physics and materials science. The term 'phase transition' refers to the thermodynamic definition by Ehrenfest [2]. It should be noted that heterogeneous transformations which can be analyzed by microscopy are first-order phase transitions, whereas transformations which do not proceed through a heterogeneous state are higher-order phase transitions and can be analyzed microscopically only if secondary features such as stress relaxation twinning of the transformation product occur.

Among the phase transitions the first-order structural phase transformations are particularly pertinent to the understanding and development of materials. The most common, diffusion controlled structural phase transformations, are associated with changes in crystal structure and composition and proceed by nucleation and growth. In some cases the crystal structure does not change (isostructural decomposition), in others the structural transformation is not accompanied by compositional changes, most noticeably in the case of diffusionless transformations. Structural

Table 1. Classification of structural phase transformations according to growth processes (after [3])

Thermally activated		Diffusionless/displacive
Short-range transport controlled	Diffusion controlled	
Polymorphic Massive	Continuous: Precipitation Isostructural decomposition Discontinuous: Precipitation Eutectoidal decomposition	Martensitic Antiferromagnetic Ferroelectric

phase transformations can proceed by a wide variety of mechanisms which, in turn, are associated with different rate-controlling processes such that the characterizing features of each transformation are essentially its *mechanism* and its *kinetics*.

Attempts to classify structural phase transformations (e.g. [3]) are invariably hampered by the multitude of factors of influence, operative mechanisms and resulting microstructural features. It will still be useful to recollect the main types of structural phase transformations in terms of features of their growth mechanisms as compiled in Table 1 (after [3]). Special features have to be taken into account regarding, for example, order–disorder, bainitic and ω phase transformations; these, however, go beyond the scope of the present discussion.

2.3 Recent Advances

The investigation of structural phase transformations has progressed substantially along with the advances in microscopic methods as described in Vol. I of this Handbook. Particularly relevant among these are:

- High resolution methods: transmission electron microscopy (TEM), field ion microscopy (FIM), scanning tunneling microscopy (STM)/atomic force microscopy (AFM).
- In situ methods.
- Image processing and automated quantitative image analysis.

The increase in lateral resolution down to atomic dimensions and the advances in methods for localized chemical, structural, and topological analysis have been decisive in gaining new and greatly improved knowledge in the field of structural phase transformations. Moreover, improved and automated methods of structural or chemical analysis have become indispensable complementary tools for the quantitative microstructural analysis of structural phase transformations. For example:

- Field evaporation of FIM surfaces and quantitative analysis by time-of-flight mass spectrometry in an atom probe.
- X-ray and neutron small-angle scattering analysis.

- Time-resolved wide-angle conventional and synchrotron X-ray diffraction analysis of structural phase transformation on heated samples.

2.4 In Situ Observations

The quantitative investigation of structural phase transformations can be extended significantly by in situ sample treatment. Inside the electron microscope thermal and irradiation interaction processes can be realized [4]. Thus different phase transformations can be studied quantitatively both at surfaces and in thin films under well-defined conditions. Table 2 gives a summary of possible electron microscopy (EM) in situ techniques. However, it should be noted that in situ techniques, especially when applied to thin films, may lead to qualitative and quantitative modifications of structural phase transformations. Some important considerations are:

- Free surfaces can be locations for heterogeneous nucleation and may lead to an increase in the rate of nucleation.
- Free surfaces are paths of high diffusion rate and may modify the growth rate.
- Free surfaces permit stress relaxation to occur such that transformation strains may be released in a different way compared to in the bulk.
- At free surfaces, heating may lead to a number of effects which can modify the mechanism and kinetics of a phase transformation; for example, the formation of oxide and other surface layers, surface faceting and grooving, and a change of chemical composition due to evaporation or absorption of components.

These indications may suffice to suggest a careful analysis of each result obtained from thin film and free surface in situ observation regarding possible deviations from bulk behavior.

Since the kinetics of structural phase transformations are often strongly affected by the structure and morphology of the reaction front, observations of growth processes at elevated temperatures by in situ TEM investigation are a powerful tool, as will now be shown for two cases. As a first example, Fig. 1 shows a sequence of phase transformations in the system MgO/TiO_2 as observed in a high voltage TEM by Hesse et al. [5, 6]. A cross-sectional sample of a TiO_2 layer on an MgO substrate was heated up to 1400 K during

Table 2. In situ EM techniques for structural phase transformations

Technique	Parameters	Reactions
Thermal treatment	Temperature, size and shape distribution of nuclei and growing particles	Nucleation, growth and coarsening processes
Sample irradiation by: Electron beam Ion beam Laser beam	Power density, plasma parameters, sample temperature, size distribution of crystalline and/or amorphous phases	Crystallization, amorphization, transformations of crystalline phases

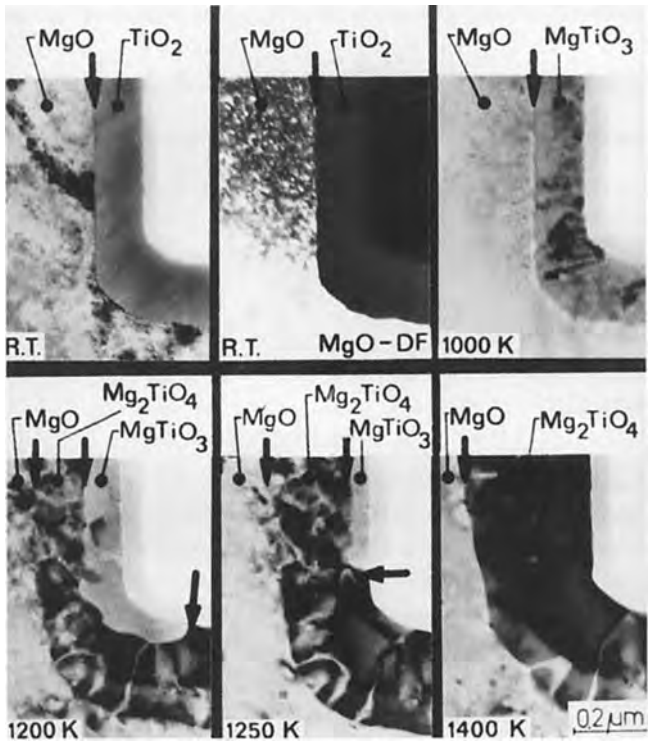


Figure 1. Phase transformation sequence in an MgO/TiO_2 cross-section by in situ annealing in a TEM [5, 6]. R.T., room temperature.

observation. At room temperature a sharp interface is visible, representing a previous cleavage surface. At about 1000 K a fine-grained $MgTiO_3$ layer forms with a high growth rate, followed by the slower formation of the more stable Mg_2TiO_4 phase at even higher temperatures. It grows at the expense of the $MgTiO_3$ layer as a completely monocrystalline phase, as concluded from in situ micro-area electron diffraction results. An interesting feature is the structure of the reaction fronts, as illustrated in Fig. 2. Both MgO and Mg_2TiO_4 crystallize in the face-centered-cubic (fcc) lattice and can form a coherent boundary. The transformation proceeds by cation diffusion, as could be shown by foreign-element marking. The unit cell of the spinel lattice of Mg_2TiO_4 ($a_0 = 0.844$ nm) has a lattice parameter almost exactly twice that of the oxygen sublattice of MgO ($a_0 = 0.421$ nm). Therefore a simple coincident site boundary $(001)_{MgO} \| (001)_{Mg_2TiO_4}$ may form as shown in Fig. 2. The position of oxygen in both lattices are marked by black dots. The cations are not shown, but different hatching indicates the different arrangements of the cations within the spinel lattice. For Mg_2TiO_4 and $MgTiO_3$ no such simple relation exists. $MgTiO_3$ has an hdp oxygen sublattice that may form a semicoherent interface with Mg_2TiO_4, corresponding to $(00.1)_{MgTiO_3} \| (111)_{Mg_2TiO_4}$.

The second example of in situ thermal treatment pertains to the crystallization of an amorphous alloy. The observation of this type of transformation by in situ thermal treatment has been applied ever since amorphous alloys have become a

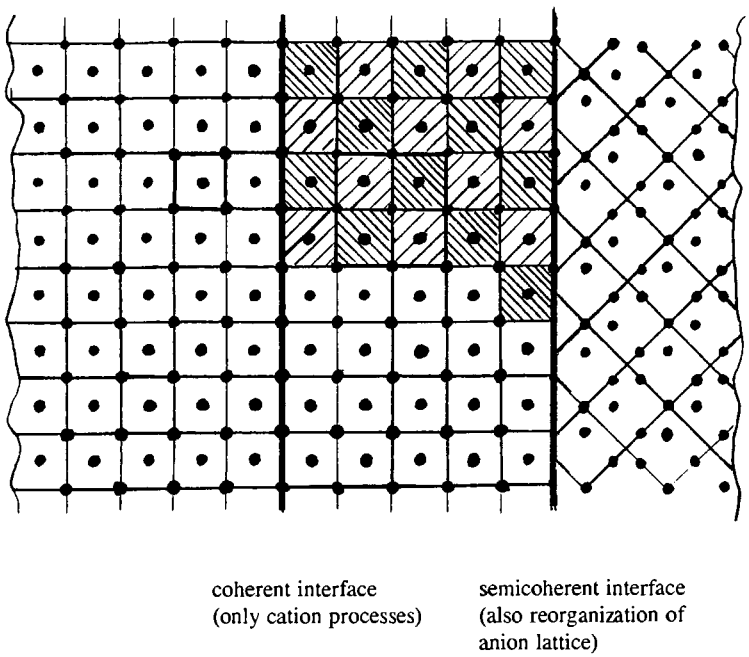


Figure 2. Structure of reaction fronts in the MgO/TiO_2 system.

subject of intensive study [7]. A direct comparison between a bright field image and a diffraction pattern is given in Fig. 3 for an in situ heating experiment of sputtered Ir–Si layers in a 200-kV TEM. Crystallization starts between 300°C and 400°C. Crystallites in the 0.1 μm size range result in ring diagrams, indicating a random orientation distribution. Above 450°C, individual rapidly growing crystals with sizes of some micrometers have formed, giving rise to single-crystal diffraction patterns. The dominant phase formed is orthorhombic IrSi; furthermore, monoclinic $IrSi_3$ precipitates.

As a second group of examples of in situ observations, crystallization and precipitation processes caused by sample irradiation, particularly with electron or laser beams will be considered. The effects are complex and depend on factors such as the type and dose of radiation, the type of sample, the temperature, and the surface state.

A particular case is the finding that the irradiation of crystalline thin layers with high-energy electrons may cause a structural phase transformation. By in situ irradiation in an analytical TEM such a transformation has been observed, as shown in Figs. 4 and 5 for a copper sample [8]. The left-hand side of Fig. 4 shows an ordered, ⟨110⟩ oriented Cu_3Au layer after low-dose ($2.3 \times 10^{27} e\,m^{-2}$) irradiation by

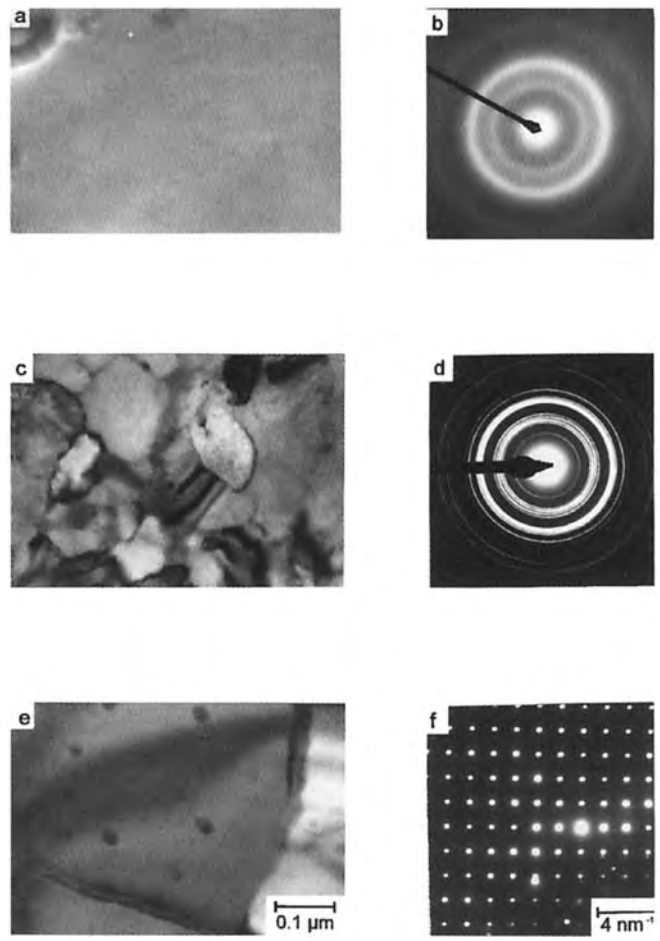


Figure 3. In situ annealing sequences of Ir–Si layers in a TEM (J. Thomas, IFW Dresden). (a, c, e) Bright field images; (b, d, f) diffraction patterns; (a, b) initial state (amorphous); (c, d) annealed up to 400°C—fine-grained, crystallite sizes about 0.2 µm; (e, f) annealed up to 700°C—coarse grained, crystallite sizes >1 µm.

400-keV electrons. The ⟨110⟩ planes of the ordered Cu_3Au phase are still visible in some places (framed). On the right-hand side of the image a transformation into a square lattice with an atomic spacing of 0.23 nm has occurred after an irradiation dose of $3.7 \times 10^{27} e\,m^{-2}$. This lattice does not correspond to any of the stable phases in the Au/Cu system. An element-analytical result was obtained by recording electron energy loss spectra, as shown in Fig. 5. The spectra were measured after electron irradiation of different doses. The net intensity at the Cu L_{23} edge decreases with increasing irradiation dose. The spectra were normalized to the Au M_{45} edge integral after background subtraction. The result was that the Cu/Au ratio in the new phase is about 1 : 1. Furthermore, the fine structure of the spectra does not show any oxidation of copper. From these results one may conclude that a new phase of the Cu/Au system has developed. Its formation may be understood from Fig. 6. With 400-keV electrons the maximum energy transfer is 19 eV to

Figure 4. TEM micrographs of ⟨110⟩ oriented Cu_3Au after 400-keV electron irradiation at total doses of $2.3 \times 10^{27} e\,m^{-2}$ (left) and $3.7 \times 10^{27} e\,m^{-2}$ (right) [8].

copper atoms and only 6 eV to gold atoms. Because the threshold energy of atom displacement in Cu_3Au is about 16–18 eV, such an atom displacement due to replacement collision sequences is possible only for copper atoms. This leads to a decrease of the copper content in the upper surface layer. When a certain concentration is reached a reconstruction occurs in this layer. As a result a new crystal structure is formed in the upper layer. By contrast, the removal of atoms from the lower specimen surface due to electron beam sputtering is possible for both kinds of atoms, because only the sublimation energy is required. This leads to thinning from the lower surface, until areas consisting of only the new phase, or even holes, occur.

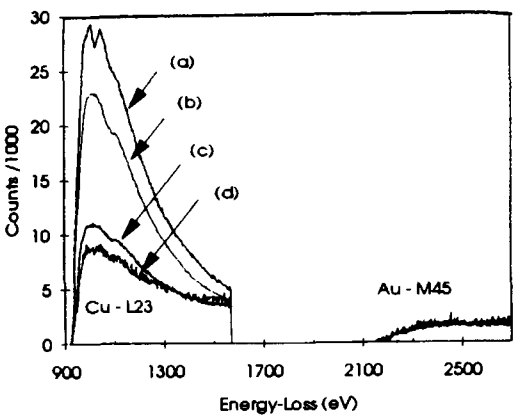


Figure 5. Electron energy loss spectra of Cu_3Au after increasing electron irradiation doses from (a) 0 to (d) $6.2 \times 10^{27} e\,m^{-2}$ [8].

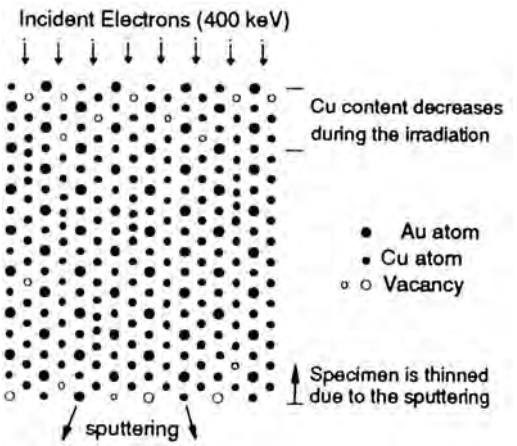


Figure 6. Formation of a new phase in a Cu/Au system [8].

The next example pertains to a phase transformation produced by laser irradiation. Figure 7 shows a transformation in a partially MgO stabilized ZrO_2 sample by in situ laser pulse irradiation in a scanning electron microscope. The sample surface was irradiated using an Nd-YAG high-power pulse laser ($\lambda = 1.06\,\mu m$; $\tau = 5\,ms$; $E'_{max} = 158\,J\,cm^{-2}$) with a Gaussian profile [9]. The depth profile of both the matrix and the laser spot was produced by means of an ion-slope cutting technique [10]. From localized X-ray diffraction patterns taken with an area detector it was

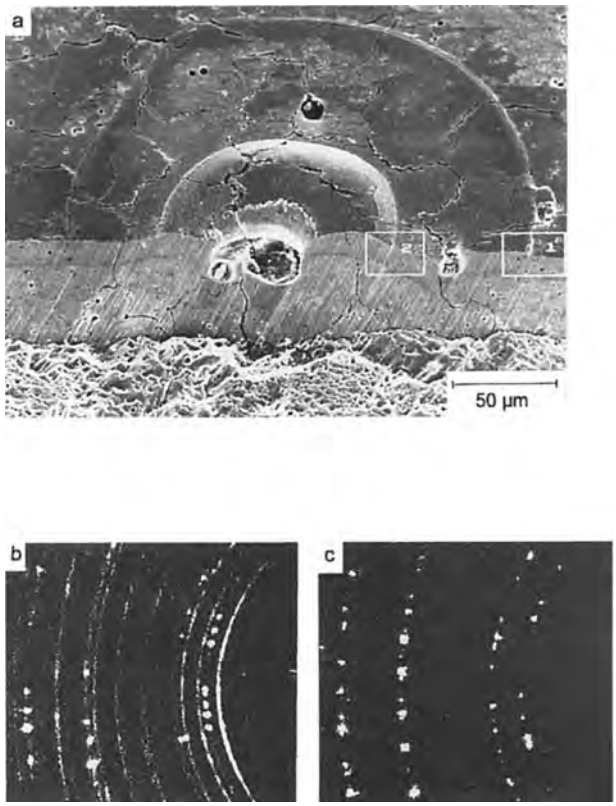


Figure 7. Phase transformation in partially MgO stabilized ZrO_2 by high power pulse laser irradiation [9, 10]: (a) ion slope cut laser spot on ZrO_2; (b) monoclinic ZrO_2 in the spot (from detail 2); (c) cubic ZrO_2 in the matrix (from detail 1).

found that the cubic ZrO_2 phase of the matrix was transformed into a monoclinic phase in the laser spot.

2.5 Investigations of Pretreated Samples

The most common microscopic investigation of structural phase transformations is carried out on pretreated samples; for example, samples annealed at a series of temperatures for different lengths of time. We also present an example of a structural phase transformation caused by ion irradiation. In the case of diffusional transformations, the microstructure observed is usually representative of the state generated by heat treatment.

However, some possible effects of specimen preparation should be kept in mind:

- Inadvertent additional 'annealing' by improper grinding, associated with heat evolution.
- 'Self-annealing' during storage at room temperature of low-melting alloys and interstitial alloys.

Such additional annealing effects may lead to additional growth, modification, or even a new variant of transformation products. In the case of diffusionless

transformations, possible effects during preparation are:

- Stress-induced formation of surface transformation products.
- Additional transformation of thin (e.g. TEM) specimens by stress relaxation effects which promote nucleation and growth [11].
- Reorientation of microstructural variants of a transformation product by stress relaxation effects.

In view of these and other possible processes which may occur in unstable or metastable states, a critical assessment of the microstructural observations is always to be recommended. The subsequent examples have been deliberately taken from a wide variety of materials and transformations.

Resistive Cr/Si films are produced by reactive magnetron sputtering in an oxygen atmosphere. In their final state they consist of conducting Cr/Si phases and isolating oxide phases. The electrical resistivity and its temperature coefficient can be adjusted by varying the oxygen content and the annealing treatment. For TEM investigations, particular thin films were deposited on NaCl; the rock salt was subsequently dissolved such that the films could be investigated by TEM without further preparation. Three examples of the structural phase transformation occurring during annealing at 600°C are shown in Fig. 8.

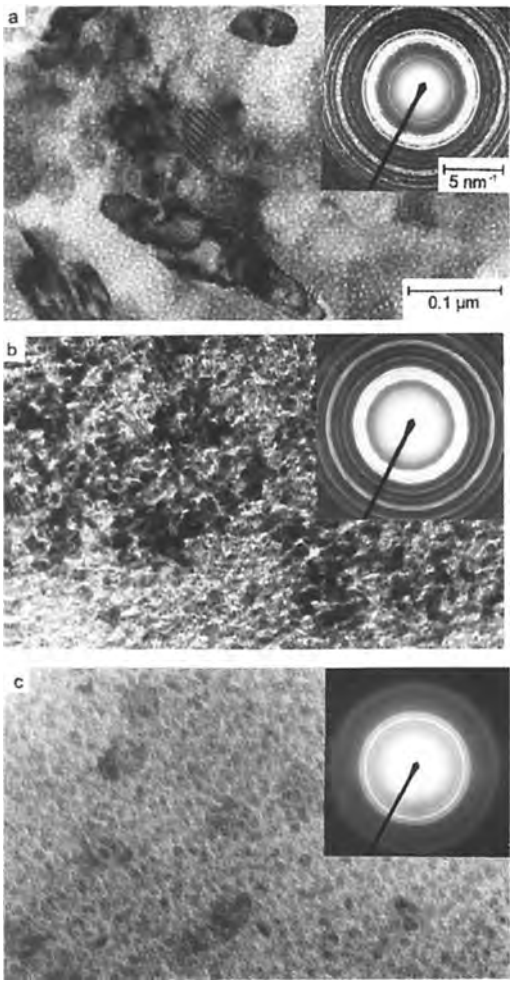


Figure 8. TEM images and electron diffraction patterns of reactively sputtered Cr/Si/O films, subsequently annealed at 600°C [12]. (a) 0 mass % O_2: $CrSi_2$ precipitates, ca. 100 nm. (b) 43 mass %: $CrSi_2$, Cr_5Si_3, and SiO_2 crystallites, ca. 20 nm. (c) 56 mass % O_2: Cr_3Si crystallites, ca. 10 nm, and amorphous SiO_2 matrix.

With increasing oxygen content, the Cr/Si phases precipitated increase in chromium content ($CrSi_2 \rightarrow Cr_5Si_3 \rightarrow Cr_3Si$), SiO_2 forms in crystalline and then amorphous structure, and the size of the precipitates decreases from 100 to 10 nm. The crystallographic analysis was corroborated by quantitative analysis of the diffracted intensities and evaluation of the radial density function derived from the intensity function [12].

We now turn to an example of advanced quantitative analysis of early stages of continuous precipitation. This

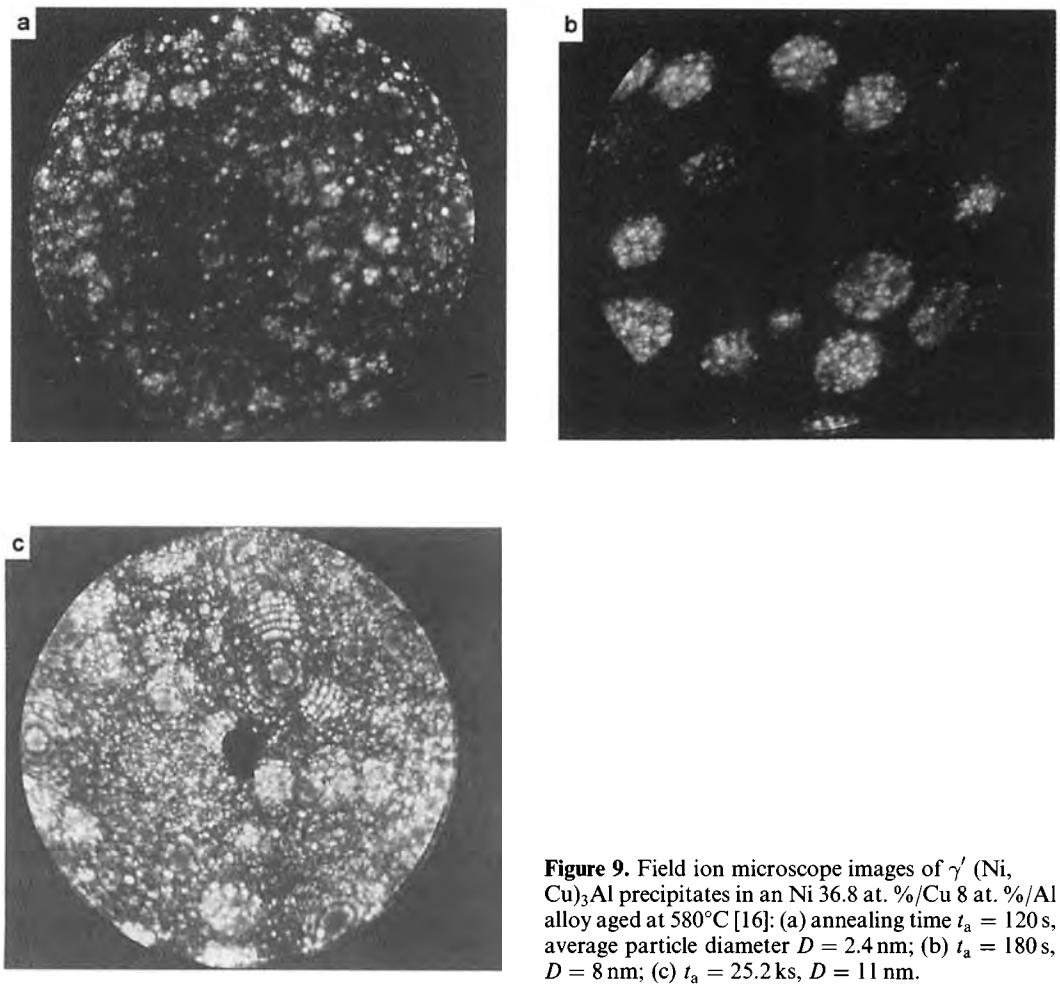


Figure 9. Field ion microscope images of γ' (Ni, Cu)$_3$Al precipitates in an Ni 36.8 at. %/Cu 8 at. %/Al alloy aged at 580°C [16]: (a) annealing time $t_a = 120$ s, average particle diameter $D = 2.4$ nm; (b) $t_a = 180$ s, $D = 8$ nm; (c) $t_a = 25.2$ ks, $D = 11$ nm.

case illustrates that atom probe, field-ion microscopy is a powerful tool. A meticulous study of coherent precipitation in Ni/Al and Ni/Cu/Al alloys was directed towards a quantitative understanding of the kinetics of precipitation from nucleation through growth till coarsening, and a comparison with theoretical treatments [13–16]. In these investigations the specimens were homogenized, quenched, and annealed in wire form at different temperatures for varying lengths of time. They were subsequently prepared into FIM tips and investigated by controlled pulsed field evaporation in combination with atom probe mass spectrometry. Typical FIM images are shown in Fig. 9, which shows progressive evolution of γ' (Ni, Cu)$_3$Al precipitates as a function of annealing time. Based on counts and analyses of field evaporated atoms, the time dependence of the average particle diameter D and the volume fraction of the precipitates, and thus the course of the transformation process, could be studied with sufficient accuracy to compare it with

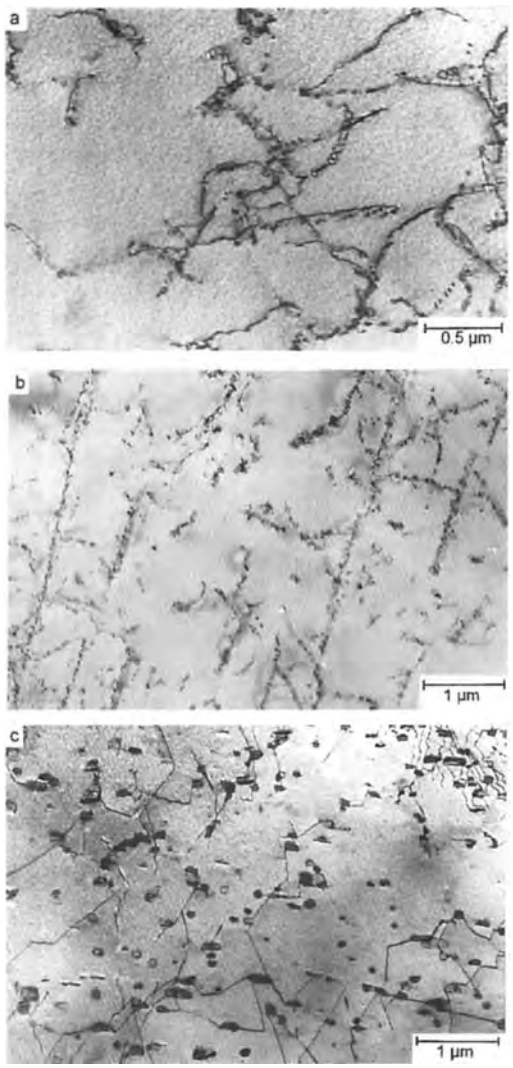


Figure 10. Vanadium nitride precipitation from α-iron solid solution in an Fe 0.1 mass %/V 0.01 mass %/N alloy via heterogeneous nucleation [17]: (a) cooled at 10 K s^{-1} from 900 to 700°C; (b) annealed for 300 s at 850°C; (c) annealed for 3.6 ks at 850°C.

the kinetics predicted by different theoretical models.

A case of heterogeneous nucleation and subsequent growth and coarsening is shown by the next example, which was based on observations of an Fe 0.1 mass %/V 0.01 mass %/N alloy [17]. Rapid cooling at 20 K s^{-1} from the α solid solution range (>900°C) to 700°C leads to the formation of dislocation loops from pre-existing dislocations (Fig. 10a). This process may be promoted by solute segregation (rather than vacancy diffusion) to the dislocations. On subsequent annealing at 850°C the loops serve as nucleation sites (Fig. 10b) by the combined effect of local solute enrichment and strain fields that reduce the strain energy per unit volume of a nucleus. Finally, prolonged annealing leads to coarsening of the particles and straightening of the dislocations (Fig. 10c). The latter observation indicates that the transient segregation effects have given way to the approach to equilibrium between the α solid solution and the vanadium nitride phase.

As a final example, transformation by ion irradiation resulting in amorphization is demonstrated in Fig. 11 [18] which shows the cross-section of a nuclear particle track caused by a 1.33 GeV uranium ion in GeS. The longer axis of the elliptic cross-section is aligned along the crystal direction with the smallest lattice parameter (i.e. with the highest density of atomic packing). The surface is parallel to (010) and irradiated perpendicularly. Whereas the lattice resolution is visible in the parent crystal, the center of the ion track has been transformed into an amorphous phase.

2.6 Concluding Remarks

High magnification and high resolution microscopic methods combined with

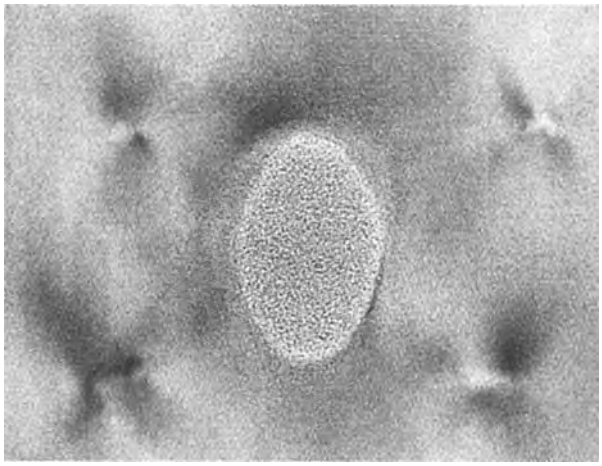


Figure 11. Amorphous track of a uranium ion in GeS, recorded using high resolution TEM [18].

diffraction and spectroscopic methods are obviously powerful tools for elucidating structural phase transformations. Advances in instrumentation and methods permitting increasingly detailed microstructural analyses for quantitative evaluation of mechanisms and kinetics, essentially down to the atomic level [19] have allowed considerable advances to be made in the understanding of transformations.

2.7 References

[1] L. Habraken, J.-L. de Brouwer, *De Ferri Metallographia*, Vol. I, Presses Académiques Européennes, Bruxelles **1966**; A. Schrader, A. Rose, *De Ferri Metallographia*, Vol. II, Verlag Stahleisen **1966**.

[2] P. Ehrenfest, *Proc. Acad. Sci. Amsterdam* **1933,** *36*, 153.

[3] H. Warlimont, *J. Microsc. Spectrosc. Electron.* **1977,** *2*, 333.

[4] K. Wetzig, *Fres. J. Anal. Chem.* **1994,** *349*, 64.

[5] D. Hesse, *J. Vac. Sci. Technol. A* **1987,** *5*, 1696.

[6] D. Hesse, J. Heydenreich, *Fres. J. Anal. Chem.* **1994,** *349*, 117.

[7] P. Furrer, H. Warlimont, *Mater. Sci. Eng.* **1977,** *28*, 127.

[8] X. Pan, W. Sigle, F. Phillipp, A. Seeger, M. G. M. Verwerft, J. Th. M. deHosson, *Proc. ICEM 13—Paris* **1994,** *2A*, 475.

[9] H. Nickel, K. Wetzig, *Fres. J. Anal. Chem.* **1994,** *349*, 36.

[10] W. Hauffe, in *In situ Scanning Electron Microscopy in Materials Research* (Ed.: K. Wetzig), Akademie, Berlin **1995,** Ch. 9.

[11] H. Warlimont, *Trans. AIME* **1961,** *221*, 1270.

[12] G. Sobe, H.-D. Bauer, A. Henke, A. Heinrich, H. Schreiber, *J. Less-Common Met.* **1991,** *169*, 331.

[13] H. Wendt, P. Haasen, *Acta Met.* **1983,** *31*, 1649.

[14] R. Kampmann, R. Wagner, in *Decomposition of Alloy: The Early Stages* (Eds.: P. Haasen, V. Gerold, R. Wagner, M. F. Ashby), Pergamon Press, Oxford, **1983,** p. 91.

[15] R. Kampmann, Th. Ebel, M. Haese, R. Wagner, *Phys. Stat. Sol. (b)* **1992,** *172*, 295.

[16] R. Wagner, R. Kampmann in *Materials Science and Technology*, Vol. 5 (Ed.: P. Haasen), VCH, Weinheim **1991**.

[17] J. Kunze, A. Güth, *Neue Hütte* **1985,** *30*, 367.

[18] R. Scholz, J. Vetters, S. Hopfe, *Rad. Eff. Def. Sol.* **1993,** *126*, 275.

[19] A. Röder, H.-D. Bauer, *Fres. J. Anal. Chem.* **1994,** *349*, 122.

Handbook of Microscopy
S. Amelinckx, D. van Dyck, J. van Landuyt, G. van Tendeloo

3 Preparation Techniques for Transmission Electron Microscopy

3.1 Indirect Investigation of Materials

3.1.1 Surface Investigation in Transmission Electron Microscopy: Replica Techniques

The geometry of the growth surface of thin films, powder materials, growth or fracture surfaces of bulk crystals, effects of different actions on the surface morphology (i.e. ion bombardment, thermal evaporation, dissolution, etc.) provide plenty of information on the material under investigation. Surface structure details below 0.1 μm are inaccessible to optical microscopy, and even with scanning electron microscopy it is difficult to resolve fine detail at sufficient contrast of the surface structure below this limit. A powerful complementary technique to light microscopy and scanning electron microscopy (SEM) is the transmission electron microscopy (TEM) replica technique [1–3].

The different replica techniques can be classified as follows:

(i) Single-stage replicas.
(ii) Two-stage replicas.
(iii) Diffuse (covering) replicas of C + P*t*.
(iv) Extraction replicas.

Except for technique (ii), these techniques damage the sample being investigated, usually by dissolution. Technique (ii) may preserve the sample, as only a plastic imprint is removed from its surface.

3.1.1.1 Preparation of Single-Stage Replicas

These replicas can be prepared from materials surfaces, which can be dissolved in an appropriate solvent, leaving a thin carbon layer and shadowing film intact. This is applicable for most metals and semiconductors.

The surface from which a replica is to be prepared must be clean. The most suitable surfaces are freshly prepared surfaces.

The sample needs to be evacuated to a pressure of about 10^{-4} mbar, then a thin shadowing film (Pt in most cases, but W, Ta or WO_3 can also be used) is deposited on the sample surface at an oblique incidence (Fig. 1). The more grazing the shadowing angle is, the longer are the shadows that surface irregularities form. Immediately after shadowing, C is deposited on the same surface to fix the Pt (W,

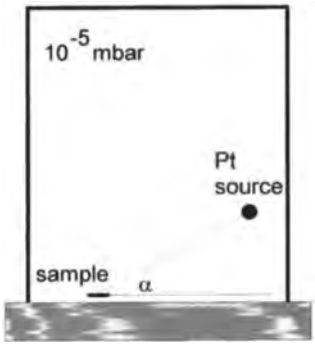


Figure 1. Geometry of Pt shadowing in the vacuum system.

Ta or WO_3) pattern formed previously. The direction of C deposition is unimportant, though it must remain at 60–90° to the surface.

In the case of a C + Pt replica, one usually uses a combined source (Fig. 2) for the deposition of both Pt and C. This source can be two carbon (graphite) rods, one of them thinned at its tip and having a hole in this tip for Pt, or Pt is screwed onto it [2, 4]. Pt is usually used in the form of a thin (e.g. 0.2 mm) wire. When heating up the carbon rod by direct heating, Pt evaporates, shadowing the sample, and then C follows, fixing the Pt distribution.

The sample–source distance is usually around 10 cm. The Pt + C replicas are then floated off in an appropriate solvent of the sample or simply in water, and the pieces collected on microgrids, usually 200 mesh, but other microgrids can also be used. The sample can be washed in distilled water a

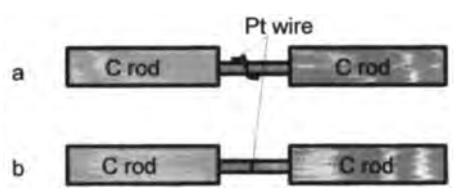


Figure 2. The source used for Pt and C evaporation.

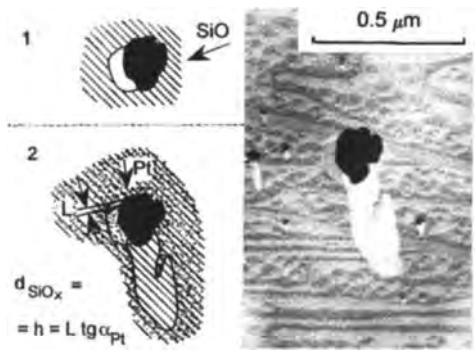


Figure 3. Measurement of height and lateral size from a double shadowed replica image.

few times before entering the transmission electron microscope, where it is viewed in the usual way. Contrast arises from uneven Pt distribution due to the shadows arising from surface irregularities. Using the length of the shadows L and the angle of deposition α, the height of the surface features h can be calculated according to

$$h = L \tan \alpha$$

Lateral sizes can be determined directly from the TEM image (Fig. 3).

The height resolution of the technique is surprisingly good: 1-nm resolution can fairly easily be achieved [5], but 0.3-nm resolution has also been demonstrated [6]. In the case of smooth metal surfaces the slip traces of dislocations can be identified (Fig. 4). Lateral resolution as good as 3 nm is achievable without the need for special procedures. Using W or Ta instead of Pt, or decreasing the temperature of the sample during shadowing and C deposition, improves the resolution of the technique in terms of both height and lateral measurements [2, 7].

The location on the sample surface can be traced fairly well, as marks on the surface or the replica make it possible to locate certain areas under TEM, to an

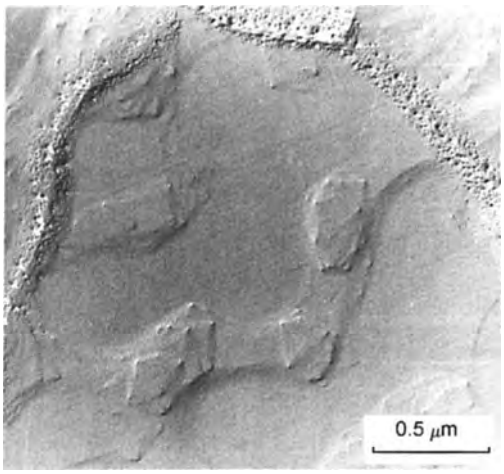


Figure 4. Slip traces of dislocations in an Al thin film seen in a single-stage C + Pt replica.

accuracy of at least 0.05 mm, or better. The internal structure of the C layer shows an 'inhomogeneity' of the order of 1 nm. This blurred contrast is not due to the structure of the surface but is inherent in the amorphous carbon thin film [8].

Characteristic thicknesses used for replica preparation are 3–5 nm for Pt and 10–50 nm for C. The steps in the preparation of a single-stage replica are shown in Fig. 5.

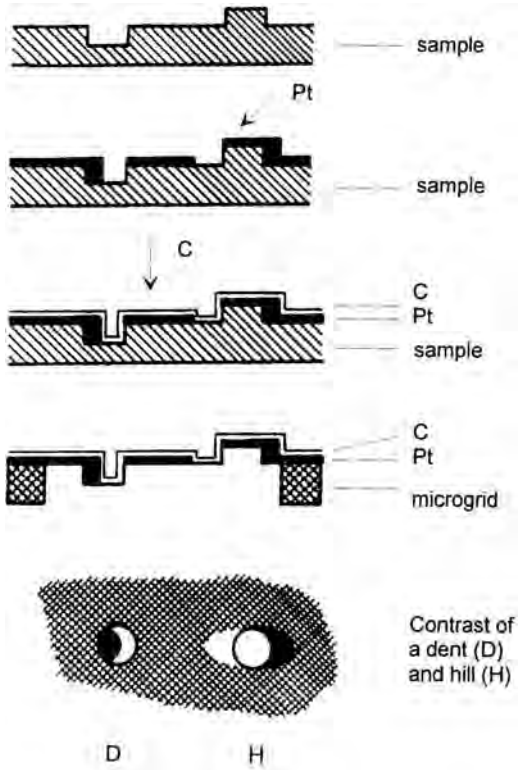


Figure 5. Steps in the preparation of a single-stage replica.

Single-Stage Oxide Replicas

Single-stage oxide replicas [1, 2, 9, 10] are prepared by oxidation of the sample surface of a metal such as Al, Ti, Ni or steels. Thermal [2, 11] or anodic [12] oxidation processes can be used. During thermal oxidation structural changes must be avoided. The oxidized surface is usually covered with an amorphous carbon layer to support the oxide film after floating off from the investigated surface, and to assure electrical conductivity in order to prevent charging in the transmission electron microscope. Nucleation of oxide formation can occur at the surface steps and at other irregularities in the surface. Thus the distribution of oxide nuclei can be used to visualize surface irregularities. The resolution of oxide replicas is close to that of C + Pt replicas.

3.1.1.2 Preparation of Two-Stage Replicas

When investigating a material which has no good solvent or which is too large to be handled in a solvent, one should prepare two-stage replicas. The technique consists of making a plastic print of the surface under investigation. The surface print so obtained is then used to make a

single-stage replica using the technique described above. The plastic print of the surface is usually made of collodion or Formvar. Collodion (cellulose nitrate) or Formvar (polyvinyl formal) [1] dissolve in organic solvents (amyl acetate and chloroform, respectively). First, droplets of a dilute solution, 1–3 vol.%, are put on the surface and allowed to dry. A more concentrated, 10–15 vol.%, solution is then applied on the top of the few thin layers formed to strengthen the print and, when dry, the print is peeled off. Drying of the thicker layers can be facilitated by using an infrared lamp. Obviously, the surface morphology shown by the two-step replica is the negative of the single-step replica (i.e. a feature that appears as a protrusion in the two-step replica is actually a dent on the sample surface). This must be taken into account when interpreting the images.

The steps involved in the preparation of a two-stage replica are shown in Fig. 6, and an example is shown in Fig. 7. The resolution is slightly poorer than for single-stage replicas, but 10 nm can be easily achieved.

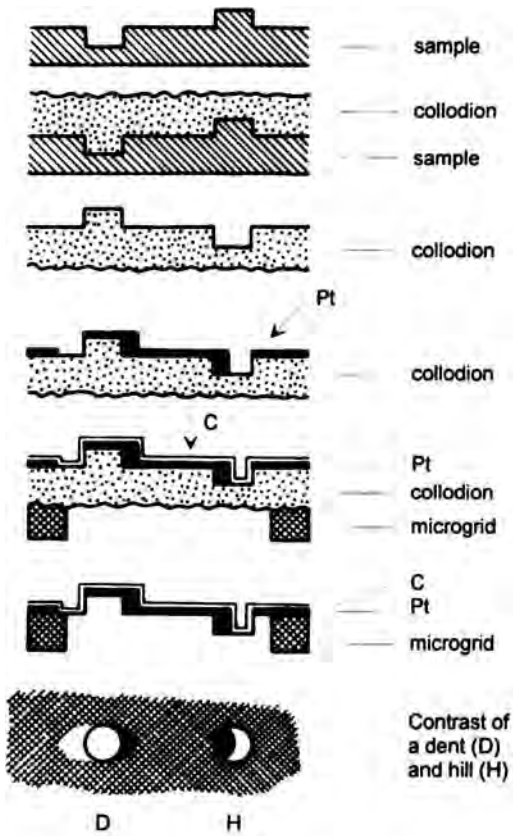


Figure 6. Steps in the preparation of a two-stage replica.

3.1.1.3 Preparation of Covering (Diffuse) Replicas

This technique is usually used when large surface irregularities are expected or when investigating powders. In essence, this is a single-step replica, but in a covering replica the direction of C deposition is important. The aim of the C deposition is to prepare a C film which covers the rough surface from each side to nearly the same thickness.

The sample is shadowed by Pt as for a single-step replica, and then C is deposited

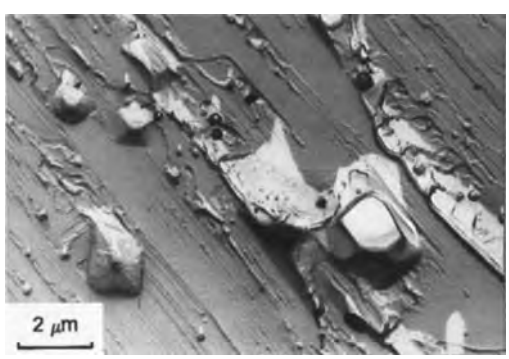


Figure 7. Example of a two-stage replica of the fracture surface of a sintered W rod displaying a rigid fracture surface and voids.

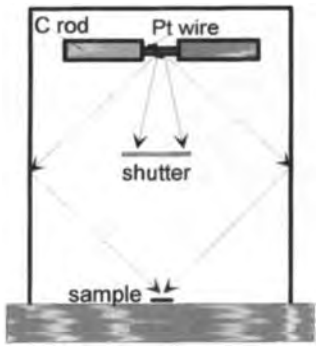
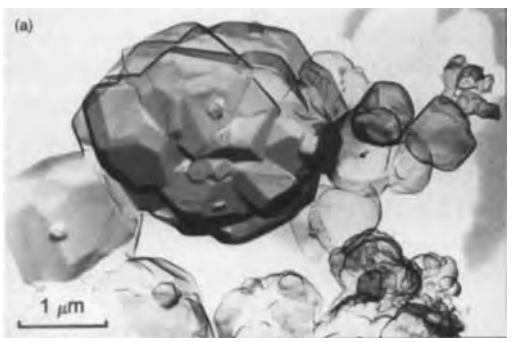


Figure 8. The geometry of C deposition for a covering (diffuse) replica.

in such a way that it cannot arrive at the sample directly from the C source. This is achieved by means of shadowing shutters (Fig. 8). In this process the very high reflectivity of the C atoms from the walls of vacuum chamber is utilized. Preferably a glass recipient is used in order to achieve higher reflectivity.

The C layer must be rather thick, 50–100 nm, so that the covering replica is still sufficiently strong after dissolving the sample. Surface irregularities with heights of around a few nanometers and powders with particle sizes of 1–3 μm can be investigated using this technique (Fig. 9). The shape and size of powder particles can also be estimated. The plastic images show the three-dimensional appearance of the surfaces of the particles, and are also useful for making stereo images in the microscope, since no diffraction contrast disturbs the stereo effect.



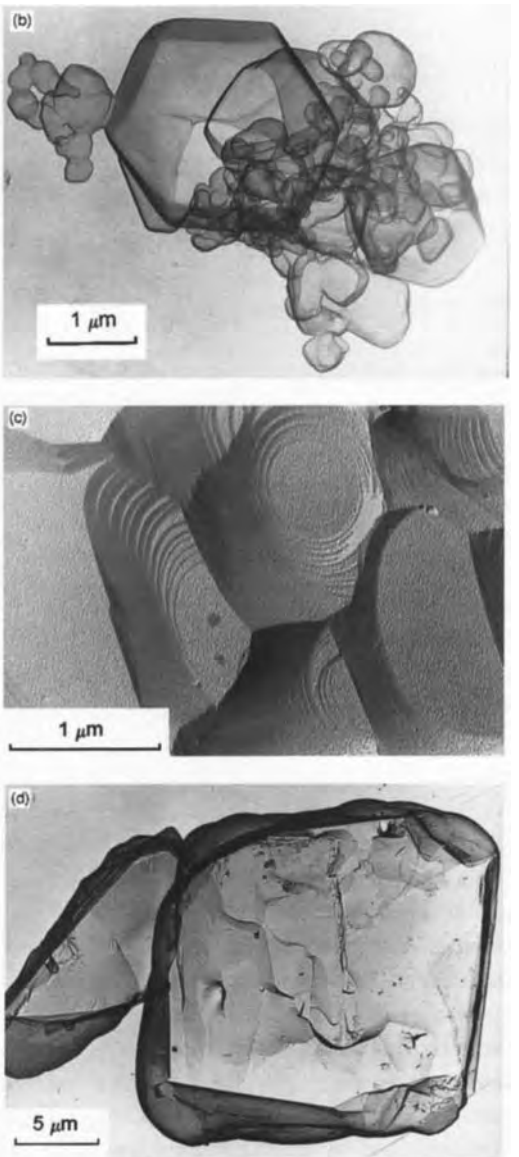


Figure 9. Diffuse carbon replica images of powdered materials: (a) W powder, shadowed with Pt; (b) W powder without Pt; (c) calcium (halo)phosphate powder used in light tubes shadowed with Pt; (d) calcium (halo)phosphate powder without Pt.

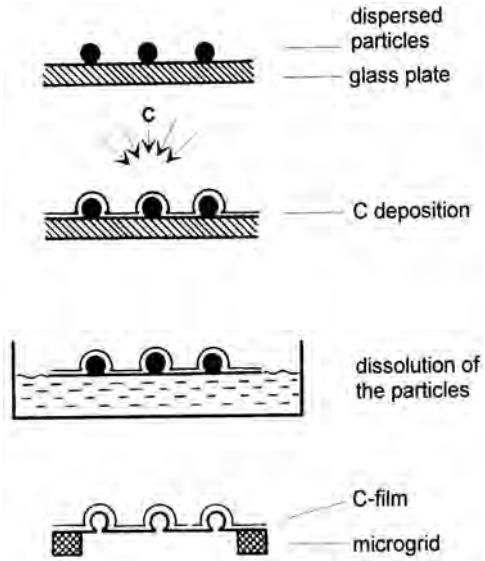


Figure 10. The steps in the preparation of a diffuse carbon replica.

The steps involved in the preparation of a diffuse replica are shown in Fig. 10.

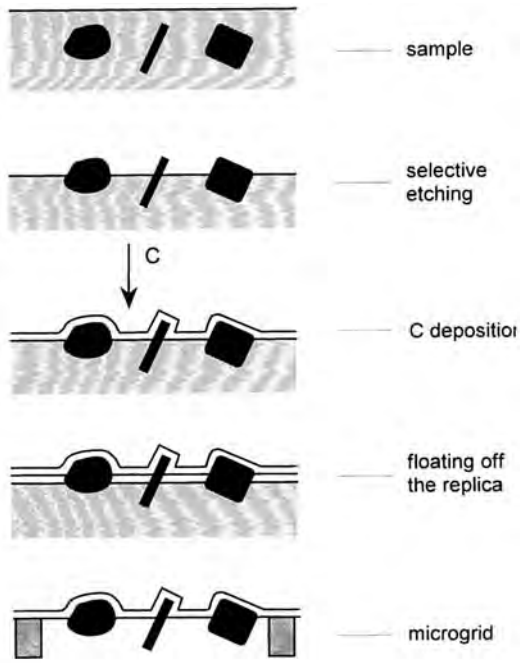


Figure 11. The steps in the preparation of an extraction replica.

3.1.1.4 Preparation of Extraction Replicas

Extraction replicas are used mainly in metallurgy when morphologies and/or the microanalysis of inclusions less than 1 μm in size is a problem [13]. The technique is illustrated in Fig. 11.

First, the two-phase structure is etched to expose the inclusion particles on the surface, and then the sample is coated with a relatively strong C film which serves as a support. No shadowing by Pt is usually used. Rotation of the specimen during C deposition or diffuse C deposition can be used to obtain a replica without discontinuities. Since the C replica follows the morphology of the etched sample surface, Pt shadowing on the top surface can be used to visualize roughly the three-dimensional shape of the inclusions.

As a further step, the replica with the supported inclusions must be floated off from the specimen by selectively dissolving the matrix material, thus releasing the inclusions and the replica. For larger areas, cutting the replica can facilitate the release process due to greater efficiency of the dissolution process. The replica can be prepared on both sides of a thin (<0.1 mm) platelet and the metal matrix dissolved through the replica films. The extracted inclusions thus obtained are sandwiched between the two C replicas, and loss of the extracted particles is eliminated or reduced [14].

The crucial task in this technique is to find a solvent which dissolves the matrix while leaving the inclusions intact. This

problem has been treated in detail for Al and Mg alloys by Carpenter et al. [15]. To avoid problems it is advisable to try different solvents, especially when investigating a new material system. The wet chemical process must also be free of particles of the resulting reaction products, e.g. oxide particles from oxidizing solvents [16]. Typical solvents for particular groups of materials are: 10% nital for steels, bromine in methanol for steels and metals, and 2% bromine in methanol at −20°C for carbide particles [17].

Non-metallic materials (e.g. ceramics or plastics consisting of granules or fibers) can be studied by removing single elements of the structure with the help of the extraction replica. Powders and other particles can also be fixed and supported in this way for investigation by TEM.

3.1.2 Decoration Technique

This technique [18] is a special case of extraction replicas where the particles fixed by the C film are usually the result of an intentionally inhomogeneous interaction of the surface with a vapor or a solution. As a result of this interaction, the active spots of the surface serve as nucleation centers for the new phase particles, thus marking the distribution of these places on the surface. The decorating material may be a metal (Au, Pt, or Ag) or another material [19]. The material is deposited on the investigated surface at an appropriate experimentally determined temperature. Surface reaction products (e.g. oxides, see above) can also be used. The same decoration effects also exist in different kinds of shadowing processes for replicas. The decoration of magnetic-flux-line distributions is also possible [20].

After the decoration process a C replica is evaporated on the decorated surface to fix the decorating particles, similar to the procedure used to make extraction replicas. The carbon film together with the decorating particles is then floated off the surface by dissolving the substrate.

The technique has been used successfully to study the cleavage and growth surfaces of ionic crystals (e.g. NaCl) [21, 22] (Fig. 12a). Thin films can also be decorated and transmitted together with the decorating particles. The technique is especially useful for studying discontinuous films and their surface topography (Fig. 12b) [23].

3.2 Direct Investigation of Materials

3.2.1 Investigation of Thin Films: Effect of the Substrate Material

Thin films are extensively investigated using TEM, partly because they are inherently prepared in a form suitable for this. The most commonly used film thickness is $<0.1\,\mu m$, and in many cases this thickness is transparent for 100–200 keV electrons. Thicker films must be thinned by one of the techniques described below.

Thin films are always prepared on a substrate. The substrate material, which influences the structure and growth of the

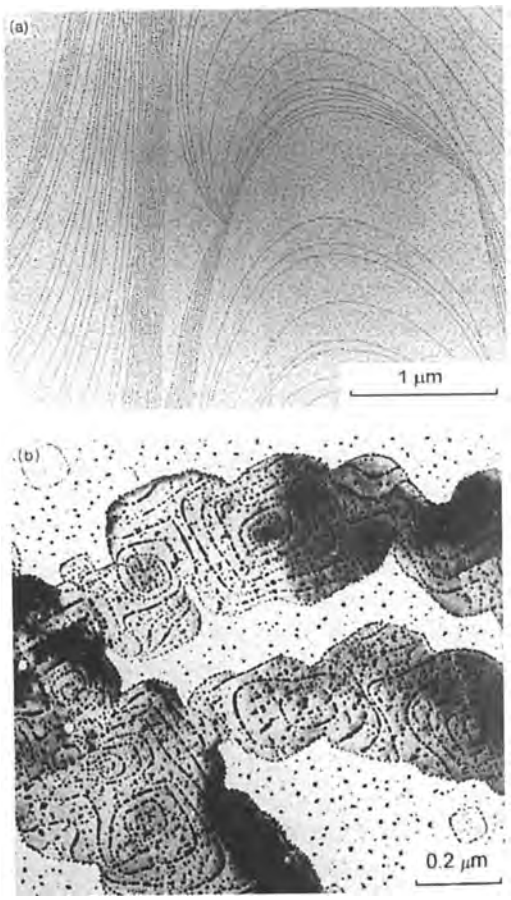


Figure 12. Examples of decorated surfaces: (a) cleavage surface of NaCl decorated with gold; (b) growth surface of SnSe crystals decorated with gold at 200°C.

film, can be thin or bulky in nature. Thin substrates are prepared intentionally for direct transmission together with the films on them. Typical of such substrates are: amorphous carbon films, amorphous SiO_x $(1 < x < 2)$ films, polycrystalline films, and cleaved single-crystal substrates made by cleaving mica or molybdenite (MoS_2).

Chemical reactions between the film and the substrate under the deposition and treatment conditions (including the thinning procedures) must be avoided, unless they are the target of investigation.

3.2.1.1 Preparation of Thin-Film Substrates

Amorphous films of C and SiO_x can be prepared on substrates from which they can easily be detached for TEM investigation. Such substrates can be, for example, the cleavage surfaces of NaCl or other ionic single crystals, or amorphous collodion or Formvar films. Their common property is that they dissolve readily in water or organic solvents. The substrate films are prepared by evaporation of C or SiO_x onto these supports; the thickness is usually 10–20 nm. The thin substrate films are then floated off their substrates in a suitable solvent. The collodion film can be annealed under vacuum at about 300°C to decompose and form a carbon layer. This C layer is usually sufficient to make the SiO_x $(1 < x < 2)$ film sufficiently conducting to prevent charge effects in the TEM. Polycrystalline supporting films of different materials can be prepared in the same way. The collodion film is prepared in a similar way to the first few layers of the two-stage replica 2 vol.% of collodion in amyl acetate is dropped onto the surface of distilled water. The collodion spreads, then the water is allowed to drain. By this method the collodion film, which then serves as substrate for the deposition of C, SiO_x or other material is caught on Cu or other (Mo or Pt) microgrids. A few tens of identical support films can be prepared simultaneously. Carbon films containing holes are prepared in the same way. The concentration and thickness of the collodion

(Formvar) when properly selected will result in hole formation.

3.2.1.2 Preparation of Cleaved-Substrate Single Crystals

This method exploits the cleavage properties of crystals, mica and MoS_2 crystals [24, 25] being the most commonly used. Thin slabs of crystal (a few mm^2) are cleaved first by using a blade, e.g. a razor blade, and then, when this is no longer possible, by using adhesive tape. For the latter, the precleaved crystals are stuck between two strips of tape, and the tapes are then separated by pulling apart the two strips. This usually separates layered crystals such as mica or MoS_2. The procedure is continued until sufficiently thin crystals are obtained, e.g. a light pink color for mica and black for MoS_2 flakes is the sign of an acceptable thickness for TEM. The flakes are then floated off from the tape in, say, acetone, washed and collected on microgrids for deposition and TEM observation.

3.2.1.3 Preparation of Thin Films on Bulk Substrates

In most cases bulk substrates need to be thinned. For this purpose the usual thinning techniques are suitable for both cross-sectional and lateral samples. As an exception, the alkali halide crystals can be considered, as they dissolve easily in water. Thin films prepared on their surfaces are usually floated off in water and collected on microgrids. The very thin or island films must be strengthened with a thin layer of C before floating.

When imaging thin films together with the substrate material, the image will always be the superposition of the images of the substrate and of the film. A structureless substrate would of course be desirable. Unfortunately, all substrate films, even amorphous films, have their own internal structure. Carbon and SiO_x films give a blurred contrast at high magnifications (of the order of 1 nm). Crystalline-substrate films give diffraction contrast of the grains and grain boundaries in polycrystalline films, while single-crystal substrates show bend and thickness contours. In the case of crystalline substrates double diffraction effects can also be important (Moiré fringes and double diffraction spots in the diffraction pattern). In addition, the defect structure of the film will appear superimposed on the defects created at the substrate–film interface.

3.2.2 Investigation of Dust Particles

The technique used to investigate dust particles depends on the purpose of the investigation and the size of the particles. When it is the surface morphology of the particles that is of interest, replica and/or scanning or tunneling microscopy techniques are used. The replica techniques are described in Sec. 3.1.1. When the object of study is the internal structure of the particles, TEM is used.

Particles less than 0.1 μm in size can be examined by TEM without further modification, if they are properly distributed on some supporting film on a microgrid. Usually, if amorphous carbon films are used, the thickness is 10–15 nm, and

these are prepared as described for the replica technique. To obtain proper distribution the powder is usually dispersed in a liquid (water, alcohol, acetone, or mixtures of these) which wets the particles, and the particles are separated by agitating the dispersion using an ultrasonic device. Droplets of the dispersion are then placed on the support film and left to dry. Nonconducting particles must be covered with a thin, diffuse-deposited C replica to ensure conductivity and fixing to the substrate film. These samples can then be investigated directly using TEM. In some cases airborne particles can also be investigated directly using TEM [26].

For particles above 0.1 μm in size, direct transmission of the particles can be a problem, and high resolution investigation is not possible. To solve this problem the particles are sectioned. The techniques most often used for sectioning are ultramicrotomy and ion milling. Crushing brittle powders in a mortar may be used to prepare particles with thin edges, and these can then be investigated directly by placing them on a support film. Ultramicrotomy of dust particles is carried out by embedding the particles in a polymerizing resin, contained within a gelatin capsule or plastic mold. Once the resin has been cured, the resin block is trimmed to a small cross-section (0.1 × 1 mm) at one end, and this is then cut into slices using a diamond blade. The slices are treated in the usual way [27, 28]. Dust particles and brittle bulk materials crushed in, for example, an agate mortar can also be investigated in this way.

Ion milling of powder materials [29–31] is carried out according to the principles of low-angle ion beam milling and sample rocking (Sec. 3.3 of this Chapter). The dust particles or powder are mixed with glue or a glue/graphite mixture so that the milling rates of the particles and the embedding material are similar. To prepare cross-sections the glue is melted and put (together with an Si single crystal about 0.2–0.3 μm thick) into the window of a Ti disc. The glue bonds and the sample is mechanically thinned in the usual way for cross-sections (to 50 μm thickness). Ion milling is carried out at angles of less than 5° to the surface. The finishing stage consists of milling with one gun from the side of the Si piece and rocking the sample about the ion beam by ±40–60° (Fig. 13a–c). This technique results in the translation of the smooth surface morphology created on the Si surface over the area of the powder (dust)/glue mixture; a smooth area about 10–15 μm wide adjacent to the Si crystal can be obtained in this way. Ion milling the sample from both sides results in fairly good sections for the TEM of powders or dust particles (Fig. 13d).

3.2.3 Chemical and Electrochemical Methods

Chemical and electrochemical methods are the most widely used techniques for preparing samples for TEM. Such methods have the advantage that preparation effects (high temperatures, deformation, and ion effects) can be avoided to a large extent compared to, for example, ion milling. With homogeneous specimens, large, thin areas can be produced. The disadvantage is that these techniques do not fit the requirements for preparing cross-sections of layered structures, except in a few cases where due to the similar

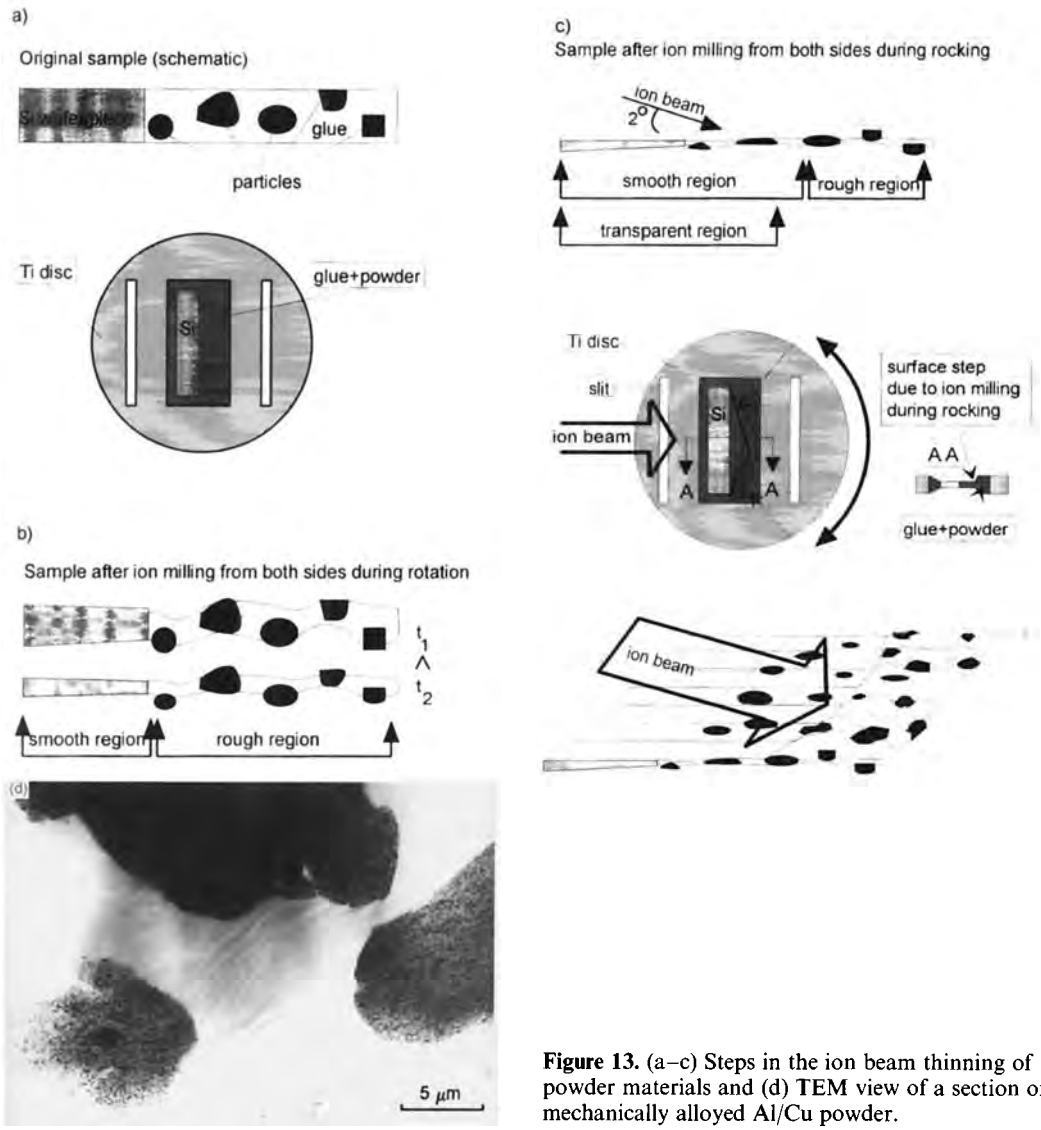


Figure 13. (a–c) Steps in the ion beam thinning of powder materials and (d) TEM view of a section of mechanically alloyed Al/Cu powder.

chemical behavior of the materials chemical cross-sectional thinning is possible, e.g. $A^{II}B^{VI}$ and $A^{III}B^{V}$ semiconductor structures [32–34].

Both chemical and electrochemical preparation follow a mechanical preparation step, which consists of cutting, mechanical thinning, and polishing. The mechanical preparation can cause damage and alter the deformation structure of the materials. The usual steps of mechanical preparation are:

(i) A sample is cut out which can fit the sample holder of a TEM. This is done using an ultrasonic cutter, diamond saw, crystal wire saw, or other tools. Usually a disc 3 mm in diameter is

formed, a size which fits the sample holder of most TEMs.

(ii) The sample is mechanically thinned to 50–200 μm thickness, generally by using SiC paper of different grades. Grinding or polishing machines can be used, or the process can be carried out by hand. Instead of thinning, deformable samples may be rolled, but this process changes the defect structure. For conducting materials, spark erosion can be used. Heating of semiconducting samples, for example by using a heated oil [35], can facilitate cutting. An advantage of spark cutting is the relatively small deformation depth, which is limited to the melted volume.

(iii) Before chemical or electrochemical treatment, the sample must be polished using diamond paste or other polishing materials. The thickness left for chemical thinning depends on the aim of preparation. If it is the deformation structure that is under investigation, the sample thickness must exceed $2d$, where d is the deformation depth of mechanical thinning. If only a phase analysis is being undertaken the sample thickness can be smaller than $2d$.

Some typical values of the deformation depth d are listed in Table 1 [36]; the values are orientational. The grain size of a polycrystalline material can provide importantinformation, since grain boundaries effectively block the dislocation movements. Thus d must always be greater than D, the grain size.

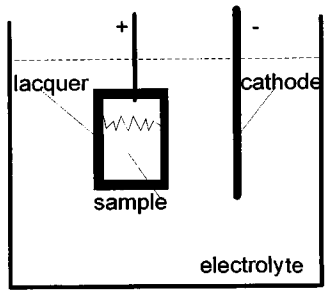


Figure 14. Outline of the window technique.

A mechanically prethinned (50–200 μm) sample can then undergo chemical or electrochemical thinning. For both chemical and electrochemical thinning there are two basic techniques [1, 3, 37–39]: the disc technique [40], for which the above-described mechanical preparation for the most part applies; and the window technique [37, 41] (Fig. 14) which applies mostly to materials in the form of fairly large sheets or ribbons. The steps are similar in both techniques, but follow somewhat different orders (Table 2).

In the window technique, the relatively large area sample is immersed in its solvent and is dissolved more or less homogeneously over its large area. Edges are usually protected by lacquer. Perforations and bridges occur between edges. The formation of bridges can be facilitated by protecting some areas (e.g. by lacquer painting) and initiating perforation in others. Areas transparent to TEM are

Table 1. Deformation depth (d) of different materials caused by mechanical thinning [36]

Material	W, Ta, Zr	Si	Al_2O_3	Al_2O_3	Steels
d [μm]	$\leqslant 25$	$\geqslant 25$	$\leqslant 2$	$\cong 10$	40–70
Technique	SiC(600)	SiC(600)	diamond paste	SiC(325)	spark cutting

Table 2. The steps of chemical and electrochemical thinning for window and disc techniques

	Window technique	Disc technique
1	Cut out an area about 1 cm^2 large from the sample	Cut out a disc, usually ∅ 3 mm, which fits the sample holder of the TEM
2	Mechanical thinning to 50–200 μm (not necessary)	Mechanical thinning down to 50–200 μm, determined by deformation depth and purpose of study
3	Dissolve the sample until perforation in one, some or many places chemically or electrochemically	Dissolve until first perforation occurs chemically or electrochemically
4	Wash carefully (water, alcohol)	Wash carefully (water, alcohol)
5	Cut a ∅ 3 mm area from the most perforated area	Sample is ready
6	If the sample is not strong enough, mount it on a suitable microgrid	
7	Sample is ready	

expected in the bridges. A piece of an area containing many small holes and many small bridges is cut out and mounted on a support microgrid; advantages can be obtained by using folded microgrids with the bars of the grids overlapping.

In the disc technique the best means of further thinning is to thin the middle area and leave a relatively thick and stable rim to facilitate handling, storing, and fixing in the sample holder. A thicker rim facilitates heat uptake in the microscope by conduction from the thin area. Usually one hole is made, preferably in the middle of the 3 mm diameter disc; the thin area is around the edges of this hole. Recipes for chemical and electrochemical thinning for many materials can be found in electron microscopy textbooks [38, 39, 42, 43] as well as in handbooks of chemistry. Some solvents of general use for electropolishing are listed by Rainforth and Thölén [44].

The solvents used in chemical or electrochemical thinning have two main functions: removing material of the sample relatively quickly, and polishing the surface. Anodic oxidation must be avoided. Before using any solvent, safety data should be collected (solvents can be explosive, inflammable, or poisonous). Chemical and electrochemical processes must be carried out with attention to ventilation and/or in a fume cupboard.

Besides the composition of the etching solution, the temperature of chemical and electrochemical etching is also important. By lowering the temperature, chemical processes slow down and dissolution characteristics become more favorable for polishing conditions. Too rapid dissolution of the material is usually unfavorable for sample preparation.

In the case of electrolytic polishing other parameters also influence the dissolution process: voltage, current, and material of electrodes. The current–voltage characteristics of a typical process is shown in Fig. 15. Each characteristic has a plateau, where polishing occurs. 'Etching' means formation of rough surfaces, leading to a very small transparent area and difficulties in interpretation. 'Pitting' means rapid local dissolution of the specimen, leading to holes and thick edges with

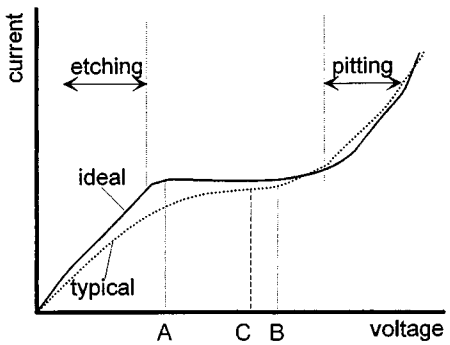


Figure 15. Typical shapes of the current–voltage characteristic of the electrochemical dissolution process.

very little or no transparent area. Decreasing the temperature usually increases the width of the plateau. The shape of the current–voltage characteristic also depends on the quantity being measured. When the voltage, rather than the sample potential, is measured the plateau becomes shorter or only an inflection point is observable on the curve. In this case electropolishing is carried out at the inflection point or a little above it [45].

There are variations of both the window and the disc techniques [1]. In the window technique, the perforation is directed to form bridges between holes and a larger thin area. For this purpose one can use local covering of (lacquer dots) or local openings on (holes in lacquer) the sample or pointed electrodes, as well as varying the electrode–sample distance during dissolution, which changes the location of the dissolution achieved with a pointed electrode. The shorter the distance to the sample, the more local the dissolution is.

A further effect which influences the dissolution processes is the movement of the electrolyte compared to a stationary one, e.g. the electrolyte can flow over the sample. This technique can be used in both the window and the disc techniques, but a more common and successful version is to use jet-electrolytic polishing. In this case the electrolyte is passed in a fine jet to one or both surfaces of the sample. Jet polishing is more commonly used in the disc technique. The main effect of the very thin jet is also to localize the dissolution effect to the area impinged by the jet. Nevertheless, protection of sample edges, using lacquers, clamps, screws, or other means, usually made from Pt or other inert metal or plastic, is necessary. The clamps or screws also fix the sample in the jet-polishing equipment. In the case of plastic (electrically isolating) fixings the electrical connection is usually made by passing a thin Pt wire through them.

A typical, although rather special example of a jet polishing device is shown in Fig. 16. This device is used to electrochemically thin W wires in both cross and lateral sections. First one side is polished, and then the other. The electrodes are made of Pt: the negative electrode is a Pt wire in the bottom chamber and the positive electrode is composed of a Pt capsule and the sample.

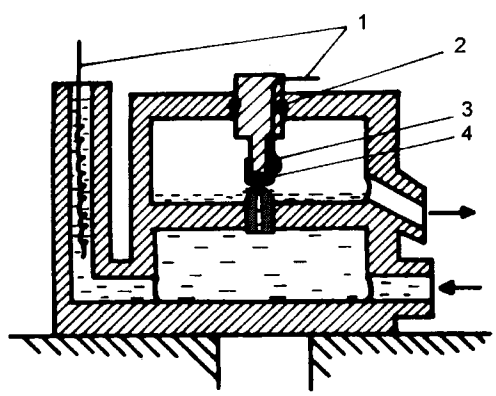


Figure 16. A jet polishing device. (1) Pt electrodes, (2) O-ring, (3) Pt cup, (4) sample.

For lateral thinning W wires (diameter $\geqslant 0.2$ mm) are ground to platelets about 0.1 mm thick, from the middle of the wire parallel to the wire axis. These platelets (3 mm long) are placed in a Ta mask (Fig. 17). The platelet and mask form a disc 3 mm in diameter, which serves as a sample for the jet polishing. The W wire is dissolved through the openings of the mask.

To prepare cross-sections, a W tip, formed by electropolishing of the wire, is pressed into a hole made in the 3 mm Ta disc. The size of the hole is $0.8h$, h being the diameter of the wire (Fig. 18). The protruding parts are dissolved electrolytically and the sample (Fig. 18c) is placed in the jet polisher.

It is very important in all polishing devices that the perforation can be detected either automatically or by using a microscope. For this purpose, the device shown in Fig. 16 is made of Plexiglas, which is completely transparent to light, and during thinning transmission observation is achieved through the whole equipment (electrolyte: 3 wt.% NaOH in water + 10 vol.% glycerin) using an optical microscope. As the above example [46] shows, there is a diverse range of electrochemical techniques, and a large number of variants on these, and special technical solutions are used in almost every individual study.

In some special cases, chemical and electrochemical thinning can be used to prepare cross-sectional samples of layered structures. The precondition is that a common solvent for both components must exist. This condition can be fulfilled in the case of semiconductor multilayer materials composed of binary or ternary systems of GaAs, AlAs, InAs, InP, HgTe,

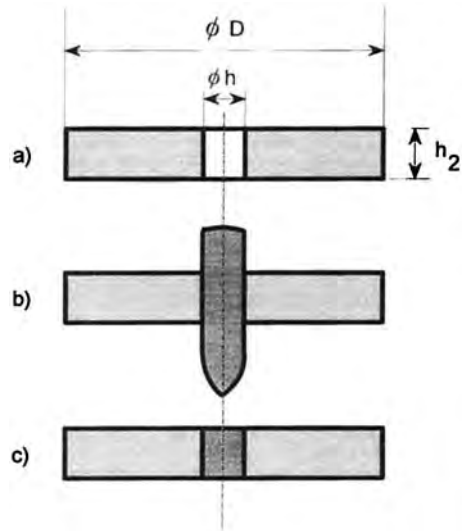


Figure 18. Schematic drawing of: (a) the Ta mask used for transverse electrochemical thinning of W wires; (b) embedding of a W wire; (c) an embedded W wire piece after dissolution of the protruding ends.

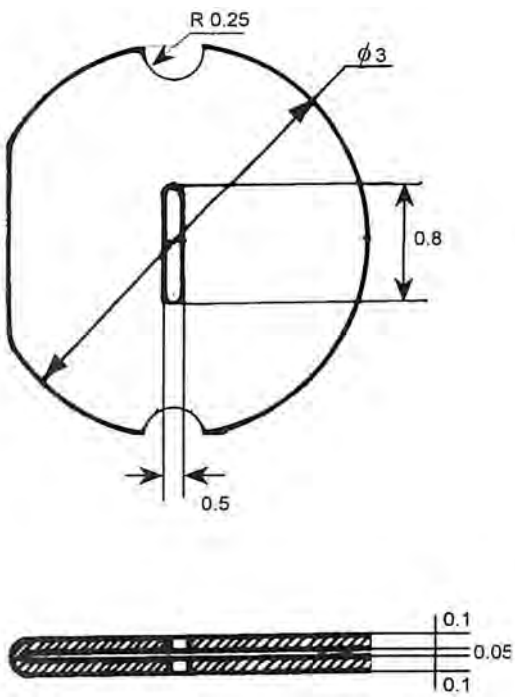


Figure 17. Tantalum mask for lateral electrochemical thinning of W wires.

or CdTe. The mechanical preparation is as usual for ion-beam etching. Two samples are glued face to face, and thinned mechanically to about 50–100 μm. Before chemical treatment the two samples are separated. The sample is placed on a glass slide with its edges protected by black vax meniscus. Chemical etching is carried out by dipping the sample in the solvent until a few microholes form, which are detected optically at a magnification of ×100 to ×200. The technique has been reported [32–34] to produce samples with measurably less damage than occurs with ion milling.

3.2.4 Ultramicrotomy

Ultramicrotomy is a technique used routinely to prepare biological samples, but it also has application in materials research [47, 48]. In particular, organic materials such as plastics, but also metals, ceramics, and composite materials, may be sectioned using this technique as an alternative to other cross-section techniques such as cleaving or ion milling.

Usually, the sample is embedded in a special epoxy such that a flat surface is exposed for cutting. A vacuum environment before hardening helps to eliminate air bubbles from the soft epoxy. The epoxy must be hard, to facilitate the preparation of thin sections and to prevent bending, which can pull the sample out of the epoxy. The epoxy usually hardens below 100°C [28].

Ultramicrotomy of the samples is carried out in two steps. First, relatively thick slabs are cut in order to trim the sample (0.1–0.2 μm thick) and then thin slabs (20–30 nm) are cut using a diamond knife. When preparing cross-sectional samples of layered structures, cutting is done either perpendicular or parallel to the interface to be examined. Brittle materials such as GaAs and Si [49] single crystals can be cut and examined under lattice resolution conditions in TEM.

The high deformation that occurs prevents the widespread use of ultramicrotomy in materials science [50]. However, the preparation of thin sections of Si and GaAs [49] thin films grown on NaCl crystals [51], surface films grown on Al [52], oxide superconductors [27], and mechanically alloyed particles [28] show that there is progress in this field.

3.2.5 Sample Preparation from Bulk Material by the Cleavage Technique

Cleavage is one of the most traditional techniques of preparing thin single-crystal samples from materials [24, 25], such as graphite, mica, and molybdenite (MoS_2), which have a layered structure. This technique has been described in Sec. 3.2.1. Recently, the need for artificial structures on single-crystal substrates or single-crystal superlattices of covalent semiconductors has revived this technique. Cleavage techniques have been developed for Si, Si/Ge, and $A^{III}B^{V}$ semiconductors; but, in principle, the techniques are applicable to all single crystals having one or more cleavage planes. All these semiconductors have a diamond-like structure and consequently cleave along the same crystallographic planes; namely, along {110} planes for (001) wafers, and along {111}

planes for (111) wafers. The angles between these planes are rather large, so the cleavage wedge available for TEM is very steep, although the sample thickness is known a priori. The steep wedge results in thickness effects in the TEM images, but the technique is quite fast and the damage caused to the sample is very small.

The cleavage technique was first established for GaAs/GaAlAs [53]. The sample wafer is polished to about 200 μm thickness and then cleaved into small squares 0.5 mm × 1 mm in size. These small chips are then mounted on special or simple slot grids [54] (Fig. 19). The tools used for cleaving and mounting are a sharp diamond pen and a pair of precision tweezers or vacuum tweezers. The same technique is used for Si, although in this case thinner wafers (~100 μm) are usually needed to give sufficiently well-cleaved corners for TEM.

By studying the cleavage properties of cubic single crystals having the diamond structure, McCaffry [55–57] developed a low-angle cleavage technique for these crystals. The essence of the technique is that the crystal is thinned to a thickness of 100 μm and scribed along a secondary cleavage plane from the rear side of the wafer. The scribing needs to be quite deep (~20 μm) to force the crystal to cleave along the plane selected: {120} for (001) orientation and {110} for (111) orientation. The crystal is then cleaved on its face side along a primary cleavage plane: {110} for a (001) and {111} for a (111) orientation. Cleavage wedges of 18.43° and 35.26° are created for (001) and (111) wafers, respectively (Fig. 20).

The cleaved chips are mounted, edge on, on a slot grid, generally by using glue [55–57], and examined using TEM.

The cleaved specimens are usually superior in quality to ion-milled ones; however, the transparent area is usually smaller for cleaved than for ion-milled specimens [56]. The cleavage technique is applicable not only to single-crystal specimens but also to thin films grown on these

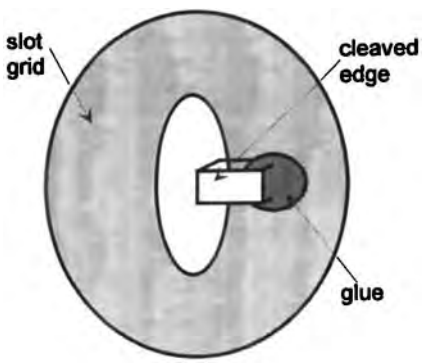


Figure 19. Mounting a cleaved sample into the slot of a simple slot grid.

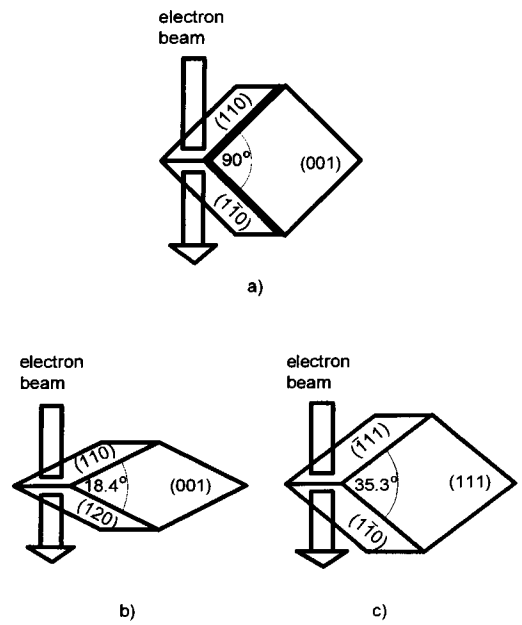


Figure 20. Schematic drawing of wedges of traditionally cleaved (a) and low-angle cleaved (001) (b) and (111) (c) crystals.

crystals, unless the films are very ductile or detach very easily from the substrate. Amorphous multilayer films have been successfully cleaved together with their Si substrate [56].

3.2.6 Artifacts due to Conventional Preparation Techniques

Different types of artifacts may arise during sample preparation, and these disturb the relevance of the information obtained from TEM. We give here a list of the most common artifacts. The thinning process is rather aggressive, and the technique used for this process must be chosen so as to exclude artifacts. In particular, unstable or metastable structures can easily change, and chemically complex materials can be damaged due to the preferential loss of one or more of their components [36, 39].

The most often observed artifacts arise from the following processes:

- Unintentional chemical reactions and unintentional chemical products.
  - (i) Preferential dissolution of components can lead to problems in overall composition in analytical TEM. The preferential dissolution of some inclusions makes problematic the determination of the morphology of the lost particles and their interfaces. Misinterpretation of missing inclusions as voids is also a problem.
  - (ii) Oxidation products covering the surface often arise during electrochemical processes. These disturb the image and diffraction, giving rise to blurred dark contrast of the image and extra rings or spots in the diffraction pattern. However, surface oxide films make possible the observation of dislocation slip traces, and can be used as oxide replicas for surface morphology studies. In the case of epitaxial oxide films, misfit dislocations may arise between the film and its oxide.
  - (iii) Chemical reaction between the reagents used in the thinning process and the sample or substrate can result in phases that represent the thinning process rather than the sample. Chemical reactions can also occur between a film and its substrate under the electron beam. Contamination with H during electropolishing (Ti or Zr) and redeposition at new locations of the once dissolved material are artificial effects which are difficult to avoid.
- Surface roughness caused by chemical reactions (etching effects) can cause problems of interpretation of the TEM images.
- The main effects of mechanical preparation are the production and movement of dislocations (see Sec. 3.2.3). Long, strait dislocations in single crystals are probably produced by preparation. The intense heating that occurs during cutting or grinding may also cause deformations or phase transitions in samples.
- Thickness effects are those arising from the fact that the sample is very thin and consequently may not represent the bulk behavior. The escape or displacement (e.g. dissociation of dislocations due to surface effects) of dislocations from thin parts of the sample is possible. Tears or cracks due to relaxation of internal

stresses can also occur. Phase transformations due to surface effects (e.g. martensitic transformation) can take place, the thickness of the film functioning as a thermodynamic parameter. Deformation twinning in thin foils (e.g. Cu) has also been observed.

- Contamination, which can be in the form of a continuous layer, spots, one or a few large flakes, or other, can occur with all types of samples prepared. Elimination by washing or ion milling may be successful in some cases.

3.3 Ion Beam Preparation

3.3.1 Practical Geometric Theory of Ion Beam Erosion: Possibility of Surface Polishing

One of the best techniques for preparing specimens for TEM investigation is ion beam thinning. The aim of TEM specimen preparation is to obtain a specimen that is thin, has a large area, is of uniform thickness and is free from artifacts. Such a specimen can be prepared by ion beam milling at low incident angles. To obtain specimens containing uniformly thick, transparent regions it is necessary to use glancing-angle ion beams, i.e. one should also avoid roughening of the surface due to ion bombardment. Reference [58] gives a detailed explanation, based on a geometrical model of changes in the surface topography during ion beam thinning, in a manner similar to those described in Refs. [59, 60]. The model provides general guidance on the practice of ion beam thinning.

When such a model is being constructed a number of simplifications can be made:

- Surface roughening by ion beams is a macroscopic process, and thus a geometrical treatment is sufficient. However, singularities of the topography, such as edges and peaks, cannot be treated in this way. On the other hand, atomically sharp edges do not develop during ion beam etching, partly because of redeposition of sputtered atoms [61], partly because of enhanced surface diffusion [62], and partly due to the transmission sputtering effect [63].
- There is no need to develop a three-dimensional treatment (the data on measured and calculated ion beam sputtering speeds of elements cover nearly one order of magnitude [64]), although it would be possible [65].

The model was developed for homogeneous materials [66–68], but by the application of statistical calculus it can be extended to describe polycrystalline materials too [69].

It is well known that almost all characteristic features of ion beam milling (sputtering speed, topographical changes, radiation damage, etc.) depend either directly or indirectly on the angle of incidence of the ion beam, θ (measured with respect to the surface normal). It is also known from the practice of ion beam thinning that both the development of a rough surface and the preferential thinning between different phases can be minimized if the angle of the ion beam is close to glancing incidence ($\theta \simeq 80$–$85°$) [66]. Topographical experiments were performed at glancing angles ($\theta > 85°$) using powerful ion guns [70] and the changes in surface topography during ion milling

were observed using a stereo light microscope. Conclusions can be drawn from the model experiments and thus the final structure of the thinned specimen can be influenced (e.g. the position of the perforation, which is important for cross-sectional thinning of layered structures).

3.3.1.1 The Geometrical Model of Surface Topography

The basic elements of the model are inclined macro steps on a smooth surface (Fig. 21). The value of the angle of inclination α is limited to the range: $0° \leqslant |\alpha| \leqslant 90°$. It is measured from the smooth region of the surface. The sign of α, which originates from the coordinate system of the sample, can be determined as shown in Fig. 21.

Further variables in Fig. 1 are t_1 and t_2 which represent a certain time interval during sputtering; H and h are step heights; $V(\theta)$ and $V(\alpha, \theta)$ are erosion speeds measured in the direction of the surface normal of the average surface and of the inclined step, respectively. θ or γ are the angles of incidence of the ion beam measured with respect to the surface normal and the surface plane, respectively.

The position of a step on the surface in time t_2 is displaced relative to the step position at time t_1. The speed of the step movement $\mathrm{Sm}(\alpha, \theta)$ along the lateral direction can be determined as follows:

$$\mathrm{Sm}(\alpha, \theta) = \frac{V(\alpha, \theta) - V(\theta)\cos\alpha}{(\cos\alpha)(\tan\alpha)} \qquad (1)$$

Equation (1) shows that the sign of $\mathrm{Sm}(\alpha, \theta)$ can be either positive or negative, as also shown in Fig. 21. The direction of a step displacement can be determined from the sign of $\mathrm{Sm}(\alpha, \theta)$.

In the case of a single step, the ion beam milling can only change the position and the shape of the step. The shape of the step is changed when the inclination angle varies along the step. The step height H remains unchanged. Changes in step height h (i.e. increase, decrease, or annihilation) can take place only by concurrence of steps during their movement. The height of a new step, formed by the concurrence of steps, cannot be greater than the sum of the step heights involved in the event.

When setting up a geometrical model describing surface evolution during ion milling, it must be taken into account that usually the sample is rotated and its surface is sputtered off. It is advantageous to define a moving coordinate system and define its movement with respect to the steady coordinate system. The following input parameters of the model should be considered: sample movement, arrangement of ion sources (Fig. 22), and the normalized sputtering speed $V(\theta)/V(0)$

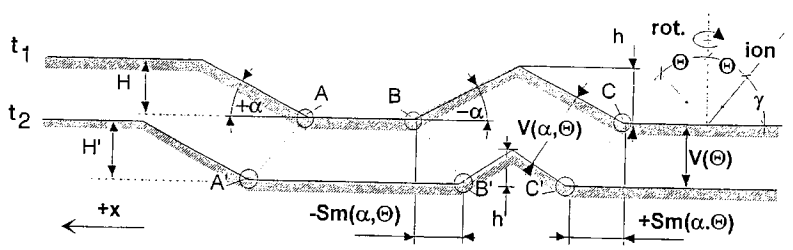


Figure 21. The basic elements of the geometrical model of surface development during ion milling.

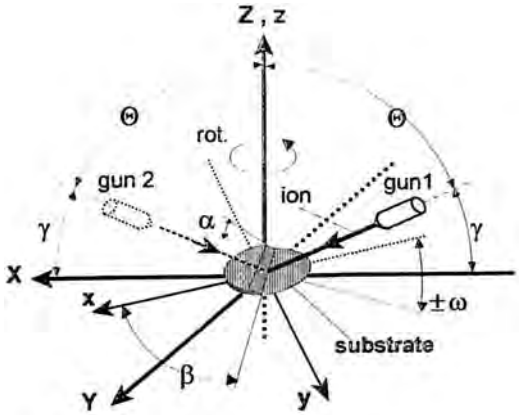


Figure 22. The experimental arrangement for thinning.

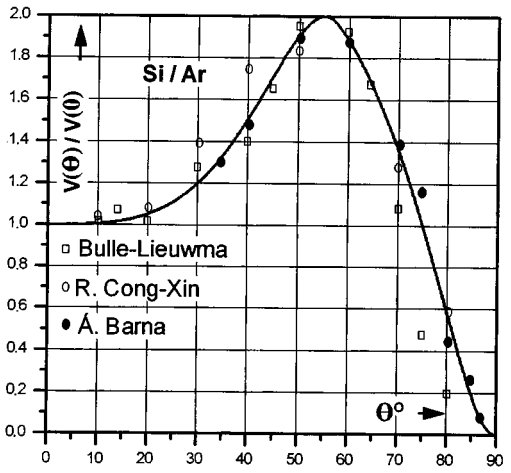


Figure 23. Sputtering speed as a function of ion beam incidence θ.

as a function of ion beam incidence angle θ (Fig. 23). The ion guns are stationary with respect to the reference coordinates (X, Y, Z) and the sample coordinates (x, y, z) are moving together with the average sample surface. All topographic changes are considered in the sample coordinate system (x, y, z).

For the experiment and calculation purposes, usually Si is used as a model material. The values of $V(\theta)/V(0)$ as a function of θ, as measured by different authors [59, 60, 71] are shown in Fig. 23. It is possible to fit an analytically defined function to the experimentally determined values (shown in Fig. 23 as a solid line).

The values of $\mathrm{Sm}(\alpha, \theta)$ need to be determined. First, $V(\alpha, \theta)$ should be computed; its value depends on the type of sample movement in its plane (static, rocked, or rotated) and on the arrangement of the ion sources.

The model does not consider the effect of ion beam shadowing because, in most cases, this is only a transient effect at the start of the thinning process. At steady state, the step angle $\alpha_{\max}$ cannot be larger than the ion beam incidence $\gamma = 90° - \theta$.

In the case of low-angle ion beam thinning ($\theta > 85°$) of a rough surface, the average sputtering speed can be more than one magnitude higher than expected from the sputtering-speed characteristic for a smooth surface (e.g. from Fig. 23), as the ion beam sees mainly the local step angles rather than the adjusted θ values in that period of thinning. Therefore, $V(\alpha, \theta)$ will be strongly dominant compared with $V(\theta)$. Ion reflection can also increase the sputtering speed of an inclined facet $V(\alpha, \theta)$, similarly to the case of cone formation [72, 73].

3.3.1.2 Rotating Sample

The thinning arrangement is shown for a rotating sample in Fig. 22. The ion sources are in mirror positions with respect to the ZY plane of the steady coordinate system. During thinning the x, y, z, coordinate system (fixed to the sample surface) is rotated with the sample and shifted along the $Z||z$ coordinates at a speed of $V(\theta)$.

When the sample is rotated the number of active ion sources is irrelevant, but their position is associated with the steady coordinate system.

Using the above notation, $V(\alpha,\theta)$ can be calculated as [60]

$$V(\alpha,\theta) = \frac{1}{2\omega_{\max}} \times \int_{-\omega_{\max}}^{+\omega_{\max}} V[\arccos(\cos\theta \cdot \cos\alpha + \sin\theta \cdot \sin\alpha \cdot \cos\omega)]\,d\omega \quad (2)$$

where ω is the angle of rotation between the reference and the sample coordinates. Equation (2) is valid for both rotation and oscillation. For rotation $\omega_{\max} = \pi$, while for oscillation $\omega_{\max}$ is arbitrary. The appropriate $\mathrm{Sm}(\alpha,\theta)$ for rotating and oscillating samples are calculated using Eq. (1). For a selected sample movement the ion beam polishing effect can be characterized by the concurrence and annihilation of inclined steps which are proportional to $\mathrm{Sm}(\alpha,\theta)$. When considering the polishing effect of the ion beam, the perpendicular sputtering speed $V(\theta)$ should be taken into account in addition to the lateral step movement speed $\mathrm{Sm}(\alpha,\theta)$.

The ratio of the lateral step movement speed $\mathrm{Sm}(\alpha,\theta)$ and the perpendicular etching speed $V(\theta)$ yields the relative step movement speed:

$$\mathrm{SR}(\alpha,\theta) = \frac{\mathrm{SR}(\alpha,\theta)}{V(\theta)} \quad (3)$$

At this point it is worth emphasizing again that the aim of TEM specimen preparation is to obtain a large, flat, electron-transparent region in the specimen. For this purpose the surface should be smooth, and this can be achieved by polishing the original surface topography during ion beam milling. Polishing is achieved when there is frequent concurrence and annihilation of steps, which occurs when the lateral step movement speed is higher than the perpendicular etching speed, that is $\mathrm{SR}(\alpha,\theta)$ is high ($\mathrm{SR} \gg 1$) for all steps (all α).

The dependence of $\mathrm{SR}(\alpha,\theta)$ on α is shown in Fig. 24. The main conclusions are:

- If $|\alpha|$ approaches zero, then $|\mathrm{SR}(\alpha,\theta)|$ approaches zero.

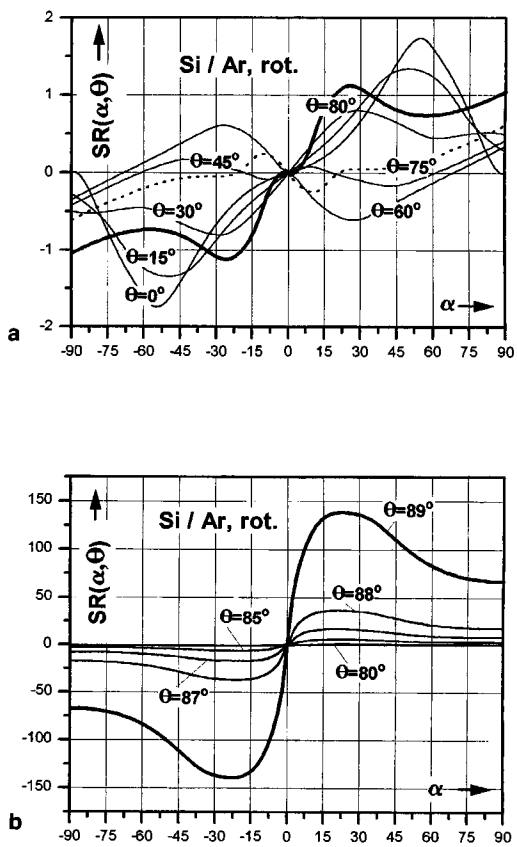


Figure 24. The dependence of the relative step movement $\mathrm{SR}(\alpha,\theta)$ on α. The angle of incidence of the ion beam θ is: (a) $\theta < 0 < 80°$; (b) $80° < \theta \leqslant 89°$.

- All SR curves are mirror symmetrical with respect to the origin of the coordinate system.
- If $\theta \geqslant 80°$ and approaches 90°, then $|SR(\alpha, \theta)|$ rapidly increases with increasing θ over the whole range of α, and its inclination at $\alpha = 0$ will be more acute.

Because of the central symmetry of the SR curves, steps for which α has opposite sign move in the opposite direction. This means that if $SR(\alpha, \theta) > 0°$ in the $+\alpha$ range then dams annihilate during the sputtering, but pits are widened with no change in depth. On continuing the sputtering, the surface exhibits narrow dams and wide pits. Finally, the dams between the pits annihilate and the whole surface becomes smooth.

3.3.1.3 Experimental Results for a Rotating Sample

When $|SR(\alpha, \theta)| \gg 1$ for $0 < |\alpha| < |\alpha_{max}|$, the surface is polished (Fig. 25a) because the lateral step movement speed is high compared with the perpendicular etching speed $V(\theta)$. When $|SR(\alpha, \theta)| \ll 1$ for $0 < |\alpha| < |\alpha_{max}|$ the original surface topography is almost completely preserved (Fig. 25b).

The ideal $SR(\alpha, \theta)$ dependence on α for ion beam polishing would be similar to a step function. Its absolute value would be large ($\gg 1$) and constant if $\alpha \neq 0$, and its sign would change at $\alpha = 0$. When θ approaches 90°, the $SR(\alpha, \theta)$ dependence on α approaches such a step function (Fig. 24b). However, if $|\alpha| < \approx 15°$ then the dependence of $|SR(\alpha, \theta)|$ on α decreases continuously as $|\alpha|$ approaches zero (Fig. 24b). The value of $|\alpha|_{max}$ decreases due to the shadowing effect during sputtering, and therefore $|SR(\alpha_{max}, \theta)|$ also decreases (Fig. 24). This means that the smaller the required $|\alpha|_{max}$ step angles to the ion-beam polished sample, the longer the sputtering time and the greater the thickness that needs to be removed from the sample in order to obtain a well-polished surface. The steepness of $SR(\alpha, \theta)$ at $\alpha = 0$ increases with increasing θ and thus the required $|\alpha|_{max}$ can be

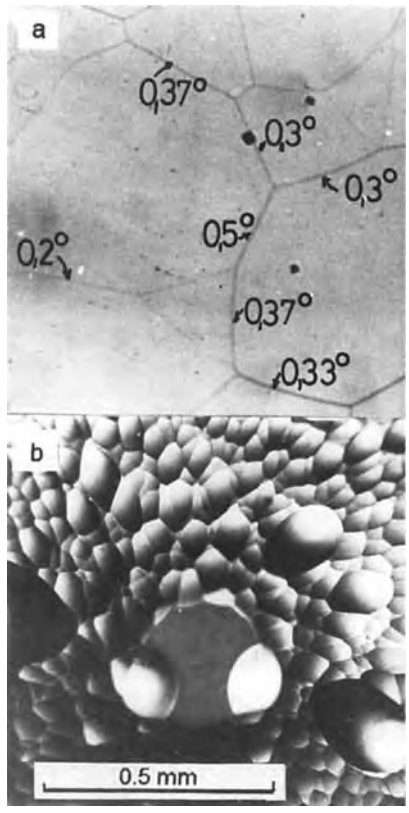


Figure 25. The surface topography of a rough Si wafer after ion-beam thinning at different angles of incidence: (a) $\theta = 88.5°$, total removed thickness 40 μm (the angles between the contacting edges of the steps were calculated from the corresponding interference micrograph; (b) $\theta = 75°$, removed thickness 280 μm.

obtained with less sample thickness being removed.

The surface of a mechanically polished sample cannot be regarded as 'homogeneous' (surface contamination and latent faults are always present due to mechanical grinding and polishing, etc.). Therefore, preferential sputtering becomes more pronounced and, in general, the surface of the sample at the beginning of sputtering becomes rough. If $|SR(\alpha,\theta)|$ is sufficiently high, then this roughness disappears; otherwise it is retained during sputtering.

To obtain a high sputtering rate $V(\theta)$, a large angle of incidence of the ion beam ($65° < \theta < 80°$; equivalent to $10° < \gamma < 25°$) is often applied. According to model calculations for these θ angles, $|SR(\alpha,\theta)|$ is low over the whole range of α. Therefore the surface topography is retained during sputtering.

If the sign of $SR(\alpha,\theta)$ is negative in the $+\alpha$ range, the direction of movement of the $\pm\alpha$ steps reverses. According to the model, pits or craters shrink but dams grow laterally. This coincides with the beam incidence range ($\theta \approx 60°$), over which the development of hummock-like surface topography is observed [59, 60, 74] for rotated samples.

It can be concluded that, when the specimen is rotated, substantial polishing, even of a rough surface, can be achieved by applying low-angle ion bombardment. It is not possible to obtain reasonably high speeds of surface-step movement for small-angle (<5°) steps.

3.3.1.4 Rocked Sample Movement

In contrast to the case of a rotating sample, both the position and number of active ion sources are important when the sample is static or rocked. As shown later, single-side milling with one ion gun should be applied to solve certain problems of cross-sectional sample preparation.

When the sample is static, bombardment using one ion gun results in translational movement of surface topography elements. This step movement results in a fluted surface structure [75] and leads to the formation of artifacts resulting in a fiber-like structure at the perforation edge.

When the sample is oscillated in its plane around the z axis at an angle of $\pm\omega$, the sputtering speed $V(\alpha,\theta,\omega,\beta)$ also increases (compared to the static case) for facets with $\beta \simeq 0$ (where β is the angle of the inclined step to the sample coordinate axis x, Fig. 22). The walls of craters or dams with $\pm\alpha$ move along the x as well as along the y direction of the sample (see the dashed curve in Fig. 26). For example, a round crater will be transformed into an elliptical one. Therefore, the possibility of step annihilation along

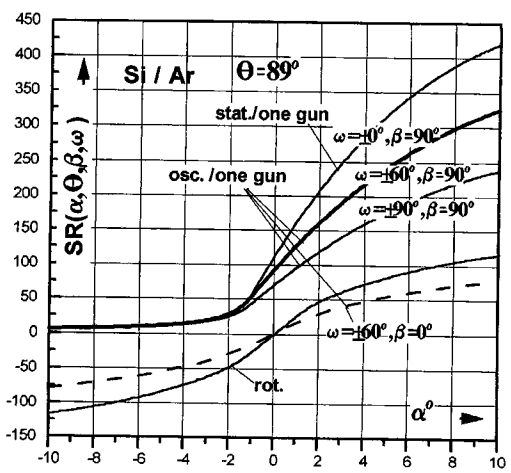


Figure 26. Dependence of $SR(\alpha,\theta,\omega)$ on α for an arrangement with one ion gun at $\theta = 89°$: ω, β, and the type of the sample movement are parameters.

the y axis is significantly increased with respect to the static configuration. The dependence of $SR(\alpha, \theta, \omega, \beta)$ on α (at $\theta = 89°$) is shown in Fig. 26 for three different sample movements (static, oscillated, and rotated) and for a thinning arrangement of one gun in the direction of x axis and steps with $\beta = 90°$ (solid lines) and with $\beta = 0°$ (dashed line). The larger ω is, the fewer artifacts develop, but also the smaller $SR(\alpha, \theta, \omega)$ is in the x direction.

The movement of a valley for one ion gun and for an oscillating sample is shown in Fig. 27. When, during the preparation of a cross-section, the sample is rotated, a valley usually develops between the two samples along the glue stripe (the area of interest) during thinning (Fig. 27a). As a result, the area of interest can easily be lost. By applying one-gun sputtering and rocking the sample, the unwanted valley structure can be moved away from the area of interest (Fig. 27b).

The dependence of $SR(\alpha, \theta, \omega)$ on α at $\theta = 89°$ for a single-gun arrangement is similar for oscillating and static samples (Fig. 28), but for the oscillating sample $SR(\alpha, \theta, \omega)$ decreases to 80% of the static-case value. Thus the oscillatory movement of the sample during ion milling preserves the advantage of static milling (high $SR(\alpha, \theta)$) but eliminates the formation of the fiber-like morphology.

The SR value can also be increased by placing two guns in mirror-image positions and oscillating the specimen, but the value is improved by only 20% compared with the rotated case (Fig. 28). This oscillatory configuration is realized by applying shading shields [76] and controlling the speed of sample rotation [77] as well. However, these techniques [76, 77] do not give rise to a translation effect, as obtained with one gun and oscillating the sample, which is useful for decreasing the effect of preferential sputtering during sample preparation.

3.3.1.5 Preferential Sputtering

The above-described model can be used to analyze how preferential sputtering influences the surface topography during cross-section thinning. Figure 29 shows a

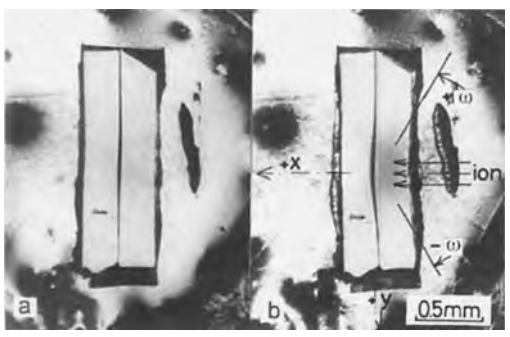


Figure 27. Optical micrographs of a cross-sectional specimen: (a) a valley at the interface; (b) the disturbing valley has been removed by applying one-gun sputtering as the sample was rocked.

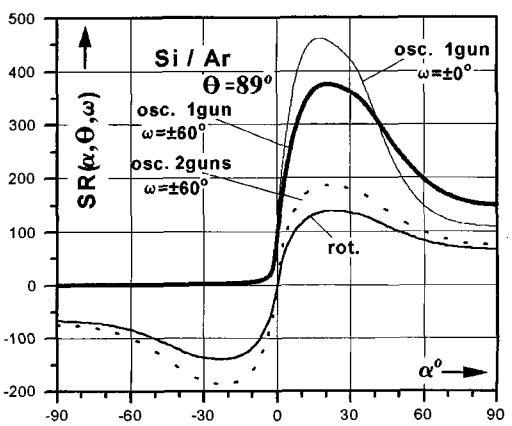


Figure 28. The dependence of SR on α for different gun and sample-movement configurations with $\theta = 89°$.

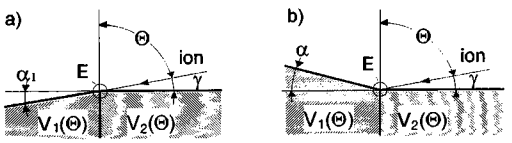


Figure 29. The effect of preferential sputtering for two different cases: (a) $V_1(\theta) > V_2(\theta)$; (b) $V_1(\theta) < V_2(\theta)$.

simplified case which can occur at the boundary of two different materials. $V_1(\theta)$ and $V_2(\theta)$ are the sputtering speeds of the two different materials, and E denotes the edge of the step which develops between them during sputtering.

In Fig. 29a, $V_1(\theta) > V_2(\theta)$ and, as a result, a step with an angle of α_1 develops at the boundary. The process is similar to the shadowing effect; α_1 varies along the step and $\alpha_{1\max} \approx \gamma$.

In Fig. 29b, $V_1(\theta) < V_2(\theta)$. In this case also an inclined step will develop at the boundary of the two different materials during sputtering. Since the edge E of the step is fixed to the boundary, one can write:

$$\tan\alpha = \frac{V_2(\theta) - V_1(\theta)}{V_1(\theta) \cdot \mathrm{SR}_1(\alpha, \theta)} \qquad (4)$$

Consequently the higher $|\mathrm{SR}_1|$ in the required α range, the smaller the value of α.

The results of the model calculation show that for one ion source and a rocked sample |SR| can be very high in the direction perpendicular to the step, even in the vicinity of $\alpha = 0°$. By applying this thinning arrangement a small value of α can be obtained, which results in a flat surface.

3.3.1.6 Retarding Field

When making TEM investigations samples should be strong enough to be handled, and thus samples should have a thick supporting rim. The maximum etching depth at the middle of the sample area is determined partly by the inner rim radius and partly by the ion beam incidence γ, and thus ion shadowing should be avoided at the center of the specimen.

When the specimen is prethinned, mechanical dimpling is frequently applied. In practice, the diameter of the dimpling wheel is 20 mm, which imposes a lower limit of ion-beam incidence (approximately 4° [78]) because of ion shadowing.

In the case of strong preferential sputtering ($V_2(\theta)/V_1(\theta) > 2$) and a thicker layer (>1 μm) on a substrate, a small angle of incidence of the ion beam ($\gamma < 1°$) is needed in order to obtain a flat surface, which contradicts the geometry of dimpling. This problem can be solved by applying a bent ion beam over the surface of the sample. When a positive (retarding) potential is applied to the sample holder the ion beam is forced to follow a parabolic path over the sample surface [78]. By adjusting the value of the retarding potential and the mechanical tilt of the ion gun it is possible to bend the ion beam in such a way that its path follows the valleys on the sample surface formed by dimpling (Fig. 30). In this case the angle of incidence of the ion beam is small and nearly uniform along the valley, in contrast to the case of the unbent beam where the angle of incidence of the beam changes along the valley.

3.3.2 Tools and Parts for Sample Preparation for Ion Beam Thinning

The steps involved in the mechanical preparation of samples for ion-beam

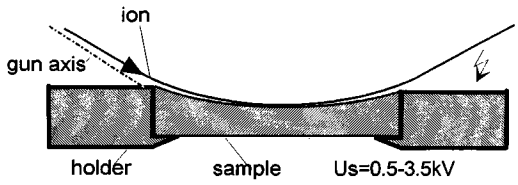


Figure 30. The effect of a retarding field.

thinning are similar to the steps in the conventional sample-preparation technique. The aim is to shape the sample so that it fits directly into the sample holder of the electron microscope (usually 3 mm in diameter). The diameter of the sample holder of the ion milling unit is also fitted to this size. The 3-mm diameter sample should not be more than a few tens of micrometers thick by starting of the ion-milling process.

The first step is to mechanically cut the area of interest of the bulk material. The cutting method used will depend on the sample material. For example, for relatively soft metals the most simple and quick method of obtaining a thin (0.5 mm) plate from the bulk material is to use a hacksaw or cut-off wheel. The minimum thickness of the plate is limited by the specimen damage that occurs during cutting, which can easily extend 200 µm below the cut surface. The entire damaged layer must be removed to avoid preparation artifacts in the sample obtained. By using a diamond slitting wheel, a wire saw, or a string acid saw, the thickness of the damaged layer can be reduced considerably.

It is possible to cut a 100 µm thick slice from Si based integrated circuits in cross-section by using a bladed diamond saw [79]; the thin slice can be used directly for ion beam thinning. A small, low speed diamond wheel saw can prove a very

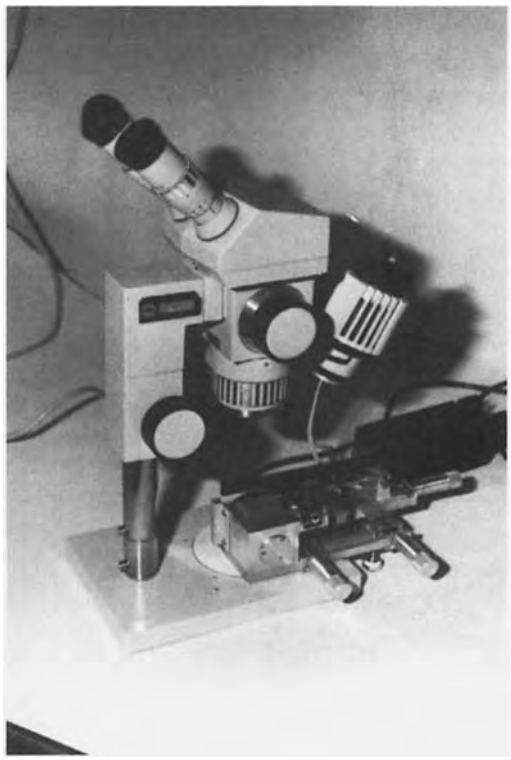


Figure 31. A low speed diamond wheel saw.

useful tool. By placing the small saw under the stereo microscope, the cutting process can be adjusted and followed closely (Fig. 31). The sample movement in x and y directions and the cutting load should be adjustable. Another basic tool used in sample fabrication is an ultrasonic or abrasive slurry disc cutter. As might be expected, this tool does not impose as much stress on the sample as does a diamond wheel saw [80]. Slurry disc cutters can be used to cut either a 3-mm diameter disc from a slice of sample material, or to cut a 3-mm diameter rod of a few millimeters length. In the latter case the necessary disc can be obtained by cutting the rod into slices. A 3-mm diameter disc can be obtained

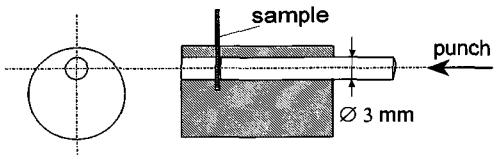


Figure 32. A simple punch tool.

from 0.03–0.5 mm thick metal plates using a disc punch. A small and very simple punch tool is shown in Fig. 32.

The next step in preparing the sample is generally mechanical grinding and a polishing process. The purpose of this step is to remove or decrease the damaged surface layer formed during cutting, and to obtain a 20–70 μm thick nearly plan-parallel disc from the thicker starting sample.

There are many commercially available disc grinders and units, but they are all more or less similar. A typical grinding and polishing tool construction is shown in Fig. 33. The sample is glued onto the central holding rod using a thin thermoplastic glue (e.g. Quartz sticky wax). The base flat of the tool is determined by the three Si_3N_4 ceramic inserts. The base plane of the tool and the axis of the sample-holder rod are precisely perpendicular to each other. The sample position relative to the base plane is adjusted using the micrometer screw.

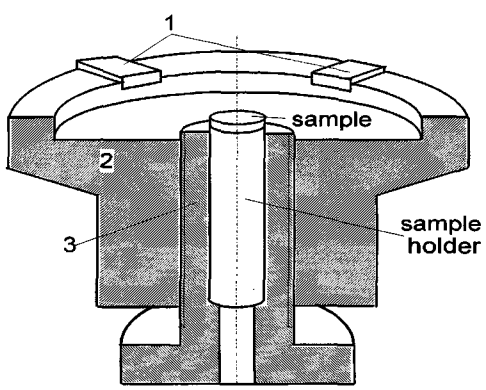


Figure 33. A typical grinding and polishing tool: 1, silicon nitride pads; 2, body part; 3, screw.

The sample is polished using first coarse and then increasingly fine abrasive paper or film until a smooth flat surface is obtained. The feeding of the sample is 10–50 μm in one grinding step. In the polishing period the sample is polished together with the grinding tool. When one side of the sample has been polished, the sample is turned over and the grinding and polishing process is repeated on the other side.

The thickness of the sample can be determined using a micrometer watch (together with the sample holder rod) or a stereomicroscope, measuring the thickness at the edge of the sample between the individual grinding steps. When the thickness of the sample is less than 60–70 μm, the second side should also be polished; generally, a nylon or perlon cloth is used. The grain size of the polishing material (alumina, diamond, etc.) is $\leqslant 1$ μm. Syton HT50 is also frequently used to polish the sample surface.

The sample can be removed from the holder rod by heating it, preferably with a small, universal heating stage under a stereo microscope. The thermoplastic glue melts and the thin sample can be pushed out of the holder rod by touching only its edge.

The above grinding and polishing process can be done by hand or by using a grinding and polishing machine. The latter is recommended when the sample material is brittle. In the case of grinding machines, the grinding load can be carefully adjusted to remain constant during the entire process.

An excellent and universal hand lapping tool is the Tripod Polisher [81]. Specific regions of Si based semiconductor devices can be thinned mechanically in cross-section to less than 0.1 μm thickness using this tool. The method needs a well developed and carefully applied lapping technique [82]. The ion beam thinning time is very short (only a few minutes). However, because of the sample preparation technique (ring technique), low-angle ion beam thinning can be applied only to one side of the sample. To avoid imposing an ion-beam shadow on the other side of the sample, the angle of incidence of the beam γ should be larger than 15–20° [83].

To reduce the ion milling time, the sample can be dimpled [80, 84–87]. This process produces a concave impression in one face (or both faces) of the sample. Dimpling leaves the rim of the sample thick, but thins the center to about 10–20 μm. The principle of the dimpling technique is shown in Fig. 34. Both the dimpling wheel and the sample are rotated; their axes are generally perpendicular to each other. Using an abrasive slurry at the contact point, a dimple is lapped or polished into the sample. The contact pressure can be adjusted. The dimpling wheel generally has a nearly spherical surface, and thus the dimpling unit must be adjusted as accurately as possible so that the axis of rotation of the sample passes through the center of the dimpling wheel. Incorrect adjustment results in a thicker area at the center of the sample [88]. During the dimpling process it is advisable to change the dimpling wheels and the abrasive slurries (from 6 μm diamond to Syton HT50) systematically. The dimpling technique can be used to thin the center of an Si sample to less than 1 μm.

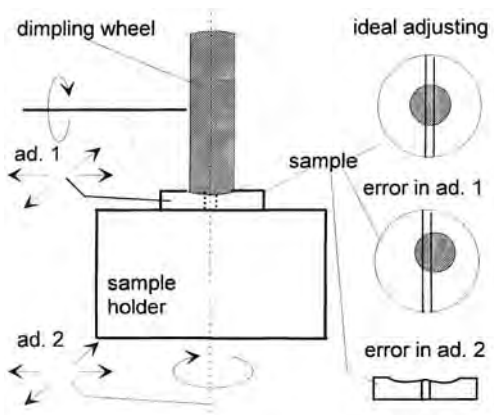


Figure 34. The principle of the dimpling technique.

The next step in sample preparation is ion beam thinning. The ion milling units are usually arranged as described by Barber [89]. Two ion guns in mirror-image positions bombard both sides of the sample simultaneously (Fig. 35). The angle of incidence of the ion beam can be adjusted either by tilting the sample holder or by tilting the common holder of the guns. During ion beam thinning the sample is rotated and the sputtering gas is let directly into the ion sources. The apparent pressure in this dynamic vacuum system is determined by the pumping speed and the gas inflow from the sources; its value is about 5×10^{-4} to 5×10^{-5} mbar. Four different types of gas ion source are in common use:

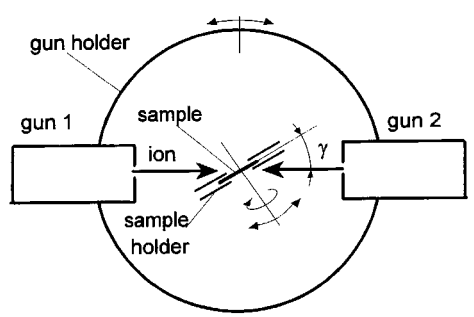


Figure 35. Ion gun arrangement for bombarding both sides of a specimen.

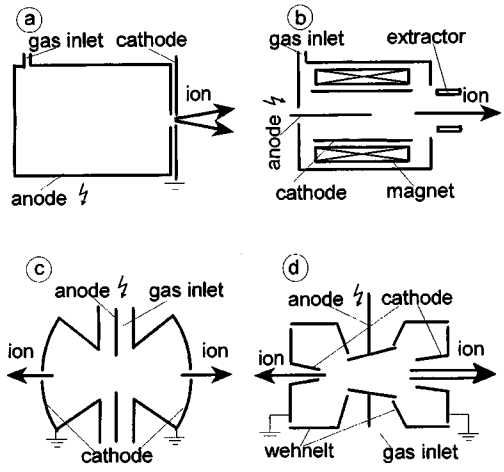


Figure 36. Ion gun constructions used for ion beam thinning: (a) hallow anode; (b) penning; (c) saddle field; (d) teletwin.

the hollow anode [90], the penning [91–93], the saddle field [94, 95], and the teletwin [90, 96] ion guns (Fig. 36).

The working pressure inside the guns is of the order of 10^{-3} mbar, except for the hollow anode gun in which the pressure is about 5×10^{-1} mbar at the anode hole. The ion beam divergence is largest for the hollow anode gun ($\sim 10°$) and smallest for the teletwin gun ($\leqslant 1°$). The anode voltage is 2–10 kV, which determines the maximum bombarding-ion energy. Due to a neutralization process that occurs near to and inside the cathode hole, a fast atom beam appears more or less beside the ion current. A fast atom gun has been developed based on a saddle field ion source [97]. In the case of the teletwin gun one can modify the fast atom/ion flux ratio from 0.1 to 0.9 by changing the front cathode from an 'ion' to a 'neutral' one.

The sputtering speed of a modern ion gun is more than $100\,\mu m\,h^{-1}$ for Si at a working distance of about 20 mm and at the optimum ($\gamma = 30°$) sputtering condition. This high sputtering speed makes possible the application of low angle ($\gamma \leqslant 10°$) thinning, in which case the sputtering speed is $5–30\,\mu m\,h^{-1}$.

3.3.3 Arrangement of Ion Milling on the Basis of the Geometric Theory

By calculating the surface step movement, an ion beam thinning arrangement can be constructed in which the necessary parameters can be adjusted and the polishing and step moving processes can be observed. Good control of the thinning can be achieved using a stereomicroscope with reflected and/or transmitted illumination. This is useful when the sample reflects the illuminating light on to the optical axis of the microscope, in which case even steps with a small height (10–50 nm) can be clearly seen. The position of the light source is more or less fixed, and thus the mirror condition is adjusted by tilting the sample holder; the basic plane during thinning is this tilted plane of the sample surface.

Optical inspection can be made of only one side of the sample, but in low angle thinning both ion guns have to work to increase the sputtering speed during polishing. Therefore, contrary to the usual arrangement, one-side thinning is used in some of the more recently commercially available ion beam thinning units (Bal-Tec, Technoorg-Linda). The position of the guns should be adjustable (when the chamber is under vacuum), especially when using high output guns, which produce ion beams of small diameter. The arrangement used for ion beam thinning is shown in Fig. 37. It is necessary to be

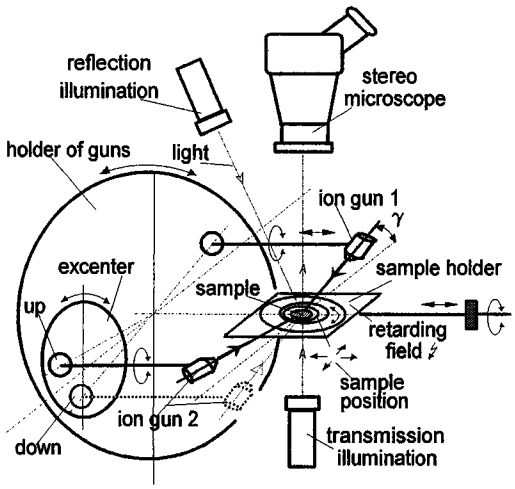


Figure 37. The ion beam arrangement in an IV3 unit with teletwin guns.

able to tilt the gun, especially when a retarding field is applied to correct the position of the bent ion beam. It is advantageous to use powerful guns operating over a wide range of accelerating voltage, for example 2–10 kV. A high voltage is needed to achieve a high sputtering rate, especially at low beam incidences, while lower accelerating voltages are needed for the preparation of sensitive materials and to reduce the formation of artifacts.

When the sample holder emits light under ion bombardment, it is useful to be able to position the ion sources under close inspection. The best solution is to use a sample holder made of Ti, which emits blue light under ion impact and has a low sputtering speed.

Besides sample rotation, a working mode in which the specimen is thinned in only one direction is needed. This can be achieved by mechanically rocking the sample or by switching the guns on and off in given sectors. In the case of sector mode operation, the dead-time is higher.

Because of the relatively small diameter of the ion beam (0.3–0.6 mm) (which is needed in low angle thinning in order to increase the sputtering speed of the sample), the area of interest of the sample should be positioned at the center of rotation of the sample holder. This is facilitated by an aperture inside the sample holder of the ion beam thinning unit. The sample holder should also be suitable for low angle thinning.

3.3.4 Ion Beam Sample Preparation for Lateral Transmission Electron Microscopy (TEM) and Cross-Sectional TEM

Ion beam thinning is one of the methods most widely used to obtain thin, transparent samples for TEM. The method is useful for all kinds of solid materials from diamond [98] to cocoon fiber [99]. The range of sputtering speeds measured for different materials is only about one order of magnitude [100] which is very small compared to the range of etching speeds in chemical or electrochemical methods.

Ion sputtering has been used to prepare TEM samples since the early 1960s. The technique almost achieves deformation-free thinning of insulating materials such as glasses, ceramics, and minerals [92, 101]. When well developed chemical methods exist for the preparation of TEM samples from metals, there is no reason to use the ion beam thinning technique because it is time consuming. However, the ion beam technique seems to be the best means of preparing TEM samples from multicomponent metal alloys containing

precipitates. Ion beam milling is generally used to prepare cross-sectional samples of layered structures.

When the surface region of a specimen or a thin layer on a substrate is to be studied, it is necessary to prepare lateral specimens. The usual method is to cut out a disc from the bulk specimen, usually using an ultrasonic disc cutter. The specimen is then ground from the rear side to the thickness required for ion beam thinning (typically 50 μm). This is followed by ion milling the specimen from the rear side until perforation. In cases when the task is to study interfaces and the layer is too thick, further bombardment from the top side is necessary. Usually, the appearance of Moiré fringes is a good sign that the interface region is in the thin section of specimen. Buried layers or regions can also be studied in plan view, but in these cases mechanical grinding should be applied from both sides to reach the appropriate region and identification of the real depth is difficult.

Many techniques have been developed for the preparation of XTEM samples. A frequently used procedure is the ring technique, the steps of which [60, 82, 102, 103] are shown in Fig. 38. The sliced pieces of the layered structures (a few millimeters long and ~2.5 mm wide) are glued together, generally face to face under pressure (Fig. 38a). A slice approximately 200 μm thick is cut from the block and ground and polished to approximately 30 μm thickness (Fig. 38b). The last step in the mechanical preparation is to glue the sample onto a holder ring (Fig. 38c). A similar technique is used to prepare samples of specific regions of semiconductor devices [83, 84]. The ring technique has some disadvantages:

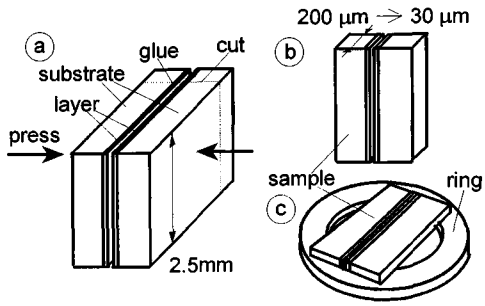


Figure 38. The steps involved in the ring technique: (a) stacking the glued samples; (b) grinding the sample; (c) gluing the sample onto a holder ring.

- Low angle ion beam thinning is possible on only one side of the sample.
- The conduction of heat between the thinned sample and the sample holder of the thinning unit is poor.
- Sputtered sample holder material can easily contaminate the area of interest of the sample.

Other frequently used fabrication procedures are the rod [80, 104] and tube [76, 105, 106] techniques. The steps involved in these procedures are shown in Fig. 39. The first step in the sample preparation is similar to that in the ring technique. A stack of samples is glued together (Fig. 39a) and a rod (3 mm diameter) cut from the stack (Fig. 39b). The sample is then ready for mechanical preparation. It does not contain any foreign material, because there is no supporting ring, and so the chance of the formation of artifacts is reduced.

In the tube technique the block of samples is embedded in a 3 mm diameter tube (Fig. 39c). The tube is usually made of metal (e.g. stainless steel), brass, or quartz. The slitted rod technique, which is a modification of the tube technique, is also applied in many cases (Fig. 39c, d). Samples for cross-sectional thinning of

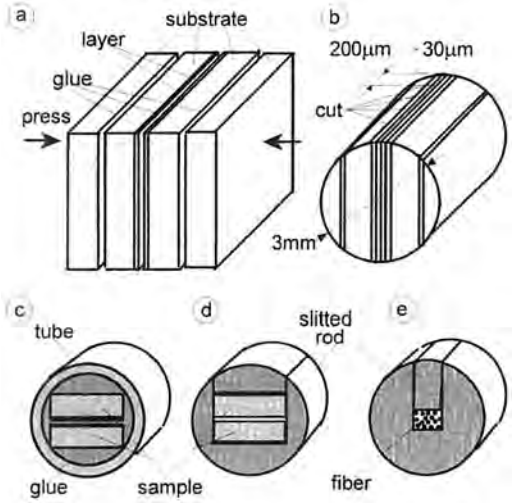


Figure 39. The steps involved in the rod technique: (a) a stack of specimens is glued together; (b) a rod is cut out and discs sliced off. (c) The tube technique. (d, e) Slitted rod technique.

thin filaments (e.g. glass) can be prepared according to Fig. 39e. The embedding process for this case is discussed later. The tube and the rod techniques avoid the disadvantages of the ring technique.

A method has been developed for the preparation of cross-sectional specimens for TEM in which the specimens are embedded. The method is rapid, relatively simple, and results in a relatively strong sample with good heat conduction and suitable for preparing specific regions. The method is particularly advantageous when low angle thinning is applied. It can be considered as a modification of the tube technique, as the basic tool used is a Ti disc (diameter 3 mm, thickness 0.3 mm) containing three slots made by electro-erosion (Fig. 40). The width of the central rectangular slot can be 0.6–1 mm and should be chosen so as to be slightly larger than the thickness of the two cut pieces to be embedded. The steps involved in this

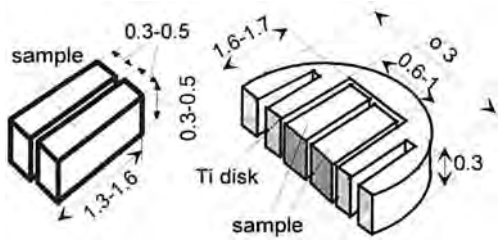


Figure 40. A section of a Ti disc used for embedding (dimensions in millimeters).

method are illustrated in Fig. 41. After cleaning, the small pieces of sample are placed face-to-face in the central slot (Fig. 41a). The walls of the central slot are then bent using a small 'edge rounded' screwdriver (hatched area in Fig. 41b) until the pieces of sample touch each other. The edges of the side slots are cut using a small sharp tool (hatched area in Fig. 41c), causing a displacement at the ends of the walls of the central slot (arrow in Fig. 41c). The walls, acting like small springs, fix the pieces of the sample. The sample is then embedded in an epoxy cement or glue; usually a one-component glue (Araldit type AT1, Ciba-Geigy) is used. The Araldit is melted on a small (15 × 15 × 5 mm) temperature-controlled hotplate under a stereomicroscope, because the whole technique requires close inspection. Melting occurs at 130–150°C and solidification begins 10–15 min later. To avoid bonding between the hotplate and the sample, a small piece of Al foil (10 × 10 mm) is placed on the hot plate. A small amount of bonding material is placed on the Al foil and, after it has melted, the sample is placed on top of the melt (Fig. 41d). As a result of the capillary force and gravity, the bonding material penetrates all the slots and fills them, appearing at the top of the Ti disc (Fig. 41e). After solidification of the bonding material (which can take place in

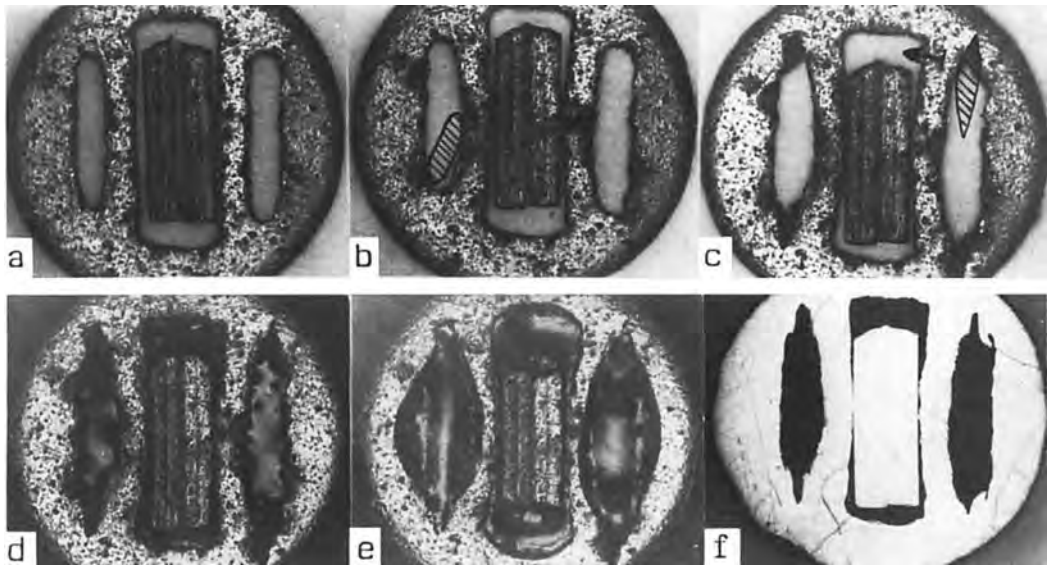


Figure 41. The steps involved in the embedding and grinding process: (a) two specimens are placed face to face into the central slot; (b) the wall of the central slot is bent; (c) the edge of the slide slot is cut; (d) the sample is placed on a drop of glue; (e) the slots and gaps are filled with glue; (f) the embedded sample is ground and polished.

a separate oven), the bonding material together with the Al foil is cut away from the Ti disc along its perimeter. This is followed by mechanical grinding and polishing on both sides of the sample to a thickness of 30–50 μm (Fig. 41f). This technique obviously has an advantage in the preparation of brittle materials due to the supporting Ti disc, without the disadvantage of the occurrence of shadowing during low angle ion beam thinning. The technique has also been used to prepare cross-sectional TEM samples of thin films deposited on water-soluble substrates [107].

The next step in the sample preparation is ion beam thinning. Here it is necessary to take into account the fact that preferential sputtering occurs at the interfaces. In accordance with Fig. 29 and Eq. (4), some simplified topographies of cross-sectional TEM samples are sketched in Fig. 42 where A, B, and C denote the substrate, layer, and glue, respectively. The samples illustrated in Figs. 42a–c were rotated and those in Figs. 42d–f were rocked during ion beam thinning.

In Fig. 42a, $V_A(\theta) > V_B(\theta) = V_C(\theta)$. The equilibrium angle [Eq. (4)] is relatively high. Perforation takes place at first in A, on both sides (development of a bridge). The heat conduction is very low. The change in sample thickness is high.

In Fig. 42b, $V_A(\theta) = V_C(\theta) > V_B(\theta)$. The situation is more favorable than in Fig. 42a. Although the equilibrium angle is the same as before, the change in thickness is reduced by a half. The probability of the first perforation is the same in A and C. The heat conduction from one side remains satisfactory.

In Fig. 42c, $V_A(\theta) = V_C(\theta) < V_B(\theta)$. In principle, an ideal plan-parallel thickness distribution occurs in B. Perforation

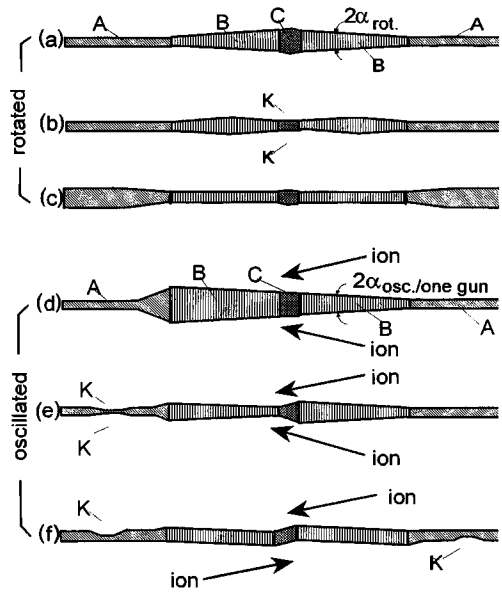


Figure 42. Simplified topographies showing preferential etching in different cases. A, the substrate; B, the layer; C, the glue. (a–c) Samples rotated at different etching rates; (d–f) samples oscillated at different sputtering rates. The sputtering rates used in (d), (e), and (f) are the same as those used in (a), (b), and (c), respectively.

occurs in B. The heat conduction is good on both sides.

In Fig. 42d, the sputtering speed ratio is the same as in Fig. 42a. The ion beam direction is the same on both sides of the sample. Because |SR| is high, the equilibrium angle is low. Despite the double-width layer, the change in thickness can be smaller than under the conditions in Fig. 42a. The first perforation takes place in one of the regions A. The heat conduction from one side is good.

In Fig. 42e, the sputtering speed ratio is the same as in Fig. 42b. The ion beam direction is the same on both sides of the sample. The equilibrium angle is low. The change in thickness is half that in Fig. 42d. The thinning process begins with the polishing of the sample surface while the sample is rotated. Therefore, a crater develops at the glue line, similarly as in Fig. 42b. During the rocking period of thinning, this crater (K) moves in the direction of the ion beam with no change in depth. At the end of thinning on the first side, the crater lies somewhere in the left side of A. A similar effect occurs during the thinning on the second side. This results in a perforation in the sample material, which is more far from the ion gun (K in Fig. 42e). The heat conduction on one side of the sample is good.

In Fig. 42f, the sputtering speed ratio is the same as in Fig. 42b. The ion beam directions are opposite on the two sides of the specimen. With high probability, a near-ideal plan-parallel thickness distribution is formed in B. Its angle of tilt from the average surface is equal to the equilibrium angle. Because of the presence of craters a bridge frequently develops, similarly as in Fig. 42a. The heat conduction is low.

In summary, two main strategies can be applied when using the rocking method. The first is to apply the ion beam in opposite directions for sputtering on the first and second sides (Fig. 42f). With this technique the thickness of the sample at the interface will be homogeneous. However, perforation might occur on both sides of the interface, creating a bridge and resulting in a poor heat conduction. Another disadvantage of this technique is that the edge of the interface is hit by the beam, resulting in rapid sputtering away of the thinnest part. The second strategy is to apply the ion beam in the same direction for both sides of the sample during rocking (Figs. 42d, f). Although in this case the thickness of the sample will

not be homogeneous, the perforation is created behind the interface, resulting in a very sharp edge. Which strategy is applied is determined according to the characteristics of the sample.

The specimen shown in Fig. 41 is shown at different stages of ion beam thinning in Fig. 43. The sample is a Ti/Cr multilayer on the oxidized surface of an Si wafer, consisting of 20 Ti layers of 13 nm and 20 Cr layers of 7 nm nominal thickness, giving a total thickness of 0.4 μm.

Figure 43a shows the sample after grinding and polishing. The interface line and the ion beams were adjusted to point into the rotation center of the sample holder, which is located in the lower quarter of Figs. 43a–d. The most favorable way to carry out polishing is to rotate the sample during ion beam thinning. The thinning should be started by polishing. For the sample in Fig. 43 this was done under the following conditions: argon sputtering gas; 10 kV accelerating voltage; two ion guns in mirror-image positions; $0.5° < \gamma < 7°$. The angle is adjusted according to the nature of the specimen; for the sample in Fig. 43 a starting value of

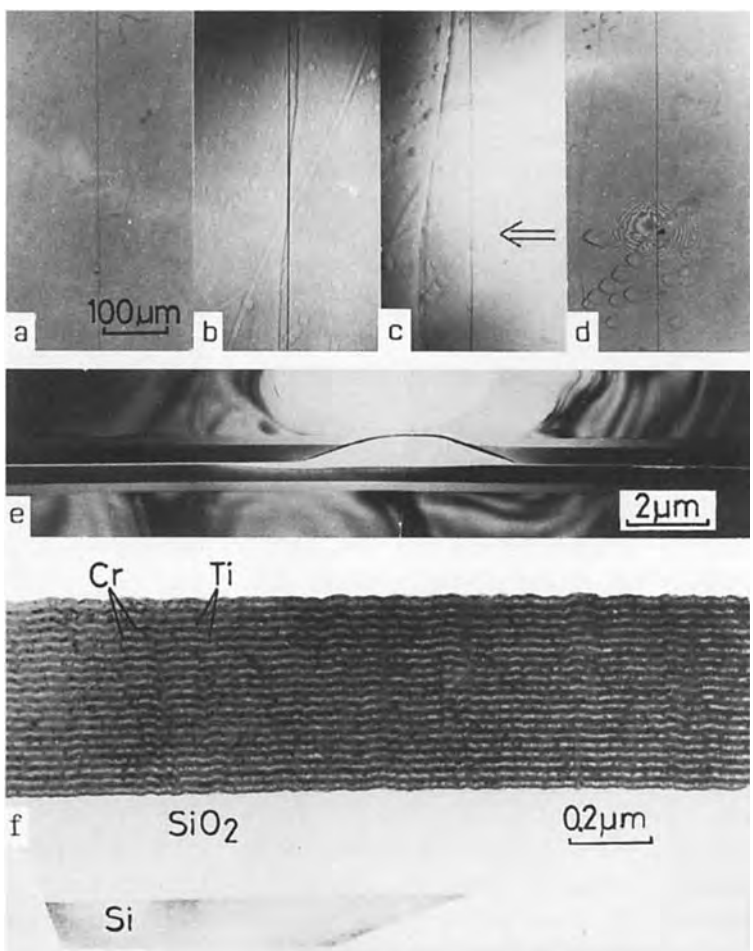


Figure 43. Images of a sample (Cr/Ti multilayer) at different stages of specimen preparation: (a) ground and polished sample; (b) sample rotated and thinned for 50 min; (c) sample rocked and thinned for a further 50 min; (d) perforated sample; (e) low magnification TEM image; and (f) high magnification TEM image.

$\gamma = 3°$ was selected. Figure 43b shows the specimen after 50 min of thinning. A valley has formed between the facing specimens. In the next step, the left ion gun was switched off, the value of γ decreased to 2° (by applying a retarding field), and the specimen rocked ($-60° < \omega < +60°$). Figure 43c shows the specimen after a further 50 min of thinning. The valley and all the small topographical elements have shifted to the left side to within about 80 μm of the center of rotation. The size and relative distance of the small craters have hardly changed compared to Fig. 43b. In this way the thinning of the first side of the specimen was stopped and the other side of the specimen then thinned in a similar way. However, in order to obtain a plan-parallel specimen, this time the right ion gun was switched off. Figure 43d shows a light micrograph of the sample after thinning, and Figs. 43e,f show 100-kV TEM images, without any correction of contrast. It can be seen that the imaged area of the sample is highly plan-parallel.

To summarize, the general method of ion beam thinning is as follows. The mechanically ground and polished specimen is ion milled from one side by using two ion guns in mirror-image positions; the angle of incidence of the ion beam is less than 5° and the sample is rotated. The wedge-shaped surface is then pushed away from the interface area by applying one gun while the sample is rocked. This process is continued until the smooth area is shifted to the interface region. (This can be controlled using the optical microscope of the ion beam milling unit.) The same procedure is repeated on the second side of the sample; in this case the rocking is continued until perforation takes place and the hole reaches the region of interest.

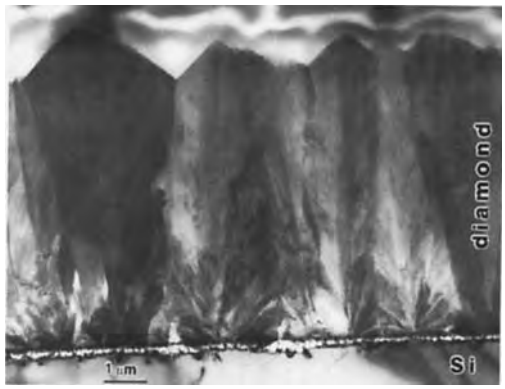


Figure 44. Cross-section of a diamond/Si sample.

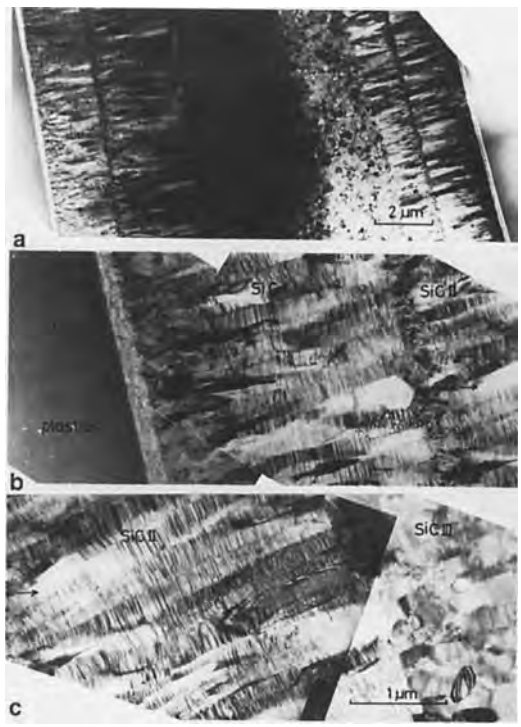


Figure 45. Bright-field images of an SiC fiber in plastic: (a) low magnification image of the whole fiber; (b) the SiC/plastic interface; (c) the middle region of the SiC fiber.

To illustrate the power of this method examples are shown in Figs. 44 and 45. The first example is a 10 μm thick diamond layer on the top of an Si substrate (Fig. 44).

The whole layer is transparent. The second example is an SiC fiber in plastic (Fig. 45), where the whole fiber, composed of different polytypes of SiC, is transparent and even the plastic/fiber interface can be investigated.

3.3.5 Artifacts Due to Ion Beam Milling

During any sample preparation method one must take into account the possible formation of artifacts, which are characteristic of the method applied. In the case of ion beam thinning the artifact-formation process is very complex. It depends on the beam energy, the beam incidence, the structure of the sample, and the chemical environment (the composition and partial pressure of residual gases). The knowledge about the radiation damage of ion-milled specimens has been collected and discussed by Barber [108]. The generation of artifacts in the surface topography (pits, cones, etc.) can be avoided by using low angle ion beam thinning.

The early stage of the ion–solid interaction has been studied extensively using computational methods [109], while fewer experimental results [110] are available. During the normal bombardment time, surface-damage usually occurs. It is well known that the depth of the ion-damaged layer decreases as the beam angle θ increases [111]. This means that, besides its other advantages mentioned in previous sections, low angle ion beam thinning decreases the amount of ion damage. To a first approximation, the thickness of the damaged layer is determined by the depth of penetration of the bombarding ions [108]. In the case of semiconducting materials, the thickness of the heavily damaged layer can be measured at the thin edge of the specimen because the original crystalline structure becomes amorphous in the heavily damaged zone. A useful way to obtain more information on the role that the energy and angle of incidence of the ion beam play in the formation of defects and ion mixing is to carry out Auget electron spectroscopy (AES) or secondary ion mass spectrometry (SIMS) depth profiling of multilayer structures having sharp interfaces.

For GaAs and Si at normal beam incidence the mass thickness of the amorphous layer is roughly proportional to the accelerating voltage E [112]. The normal component of the ion velocity is effective at oblique incidence and the thickness of the damaged layer depends only on the ion energy ($E\cos\theta$). A similar dependence has been found of the amorphous layer thickness d_a on the beam incidence θ [60]:

$$d_a = 4.5\,(E_{Ar}\,[\mathrm{keV}])^{0.5}\cos\theta\,[\mathrm{nm}] \qquad (5)$$

but d_a depends on the square root of E. Theoretical considerations regarding the mixing effect [113] result in a similar depth resolution dependence on E and θ during depth profiling; however, reliable predictions do not exist for $\theta > 75°$. The surface amorphization and the mixing effect are not the same, but they are very similarly dependent on the energy and angle of incidence of the ion beam.

When $\theta > 80°$ the amorphous-layer thickness is larger than Eq. (5) predicts [60]. An Si single crystal sample was thinned using 3-keV Ar ions at $\theta = 87°$ for 10 min; the sample was rotated during thinning. Reflection electron diffraction

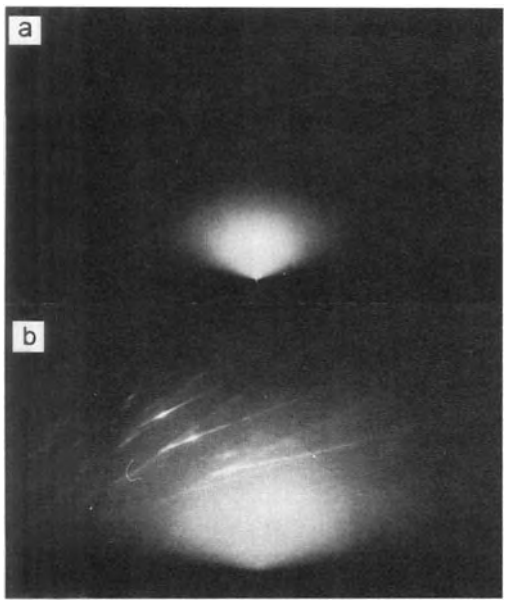


Figure 46. RED patterns (electron beam angle of incidence 1°) of an ion beam thinned Si single crystal: (a) ion energy 3 keV, $\theta = 87°$; (b) ion energy 0.5 keV, $\theta = 87°$. The sample was rotated during etching.

(RED) patterns of the thinned Si surface were taken using TEM. The angle between the sample surface and the electron beam was 1°. According to the calibration measurements, the RED pattern (Fig. 46a) showed the formation of a 1.5–2 nm thick amorphous surface layer, contrary to Eq. (5) [60] which predicts a thickness of 0.4 nm. The thinning experiment and the RED investigation were repeated for the same Si sample but with the beam energy decreased to 0.5 keV (Fig. 46b). The Kikuchi lines appeared, and the RED pattern was very similar to the pattern of a chemically cleaned Si single crystal, that is no amorphous layer was detected. The image in Fig. 46b clearly shows that it is possible to obtain a nearly damage-free thinned sample surface on elementary semiconductors by using a low-energy ion beam at a low angle of incidence.

The radiation damage to metals, which appears in the form of dark diffraction contrast over the whole sample surface, can also be reduced by using low energy, low angle ion beam thinning. Ion beam thinned W samples prepared at 3 and 0.5 keV ion energy are shown in Fig. 47. It can be seen that thinning with 0.5-keV ions significantly decreases the density of damage to the sample surface, but the damage is not as low as obtained with chemical thinning. It seems that as a secondary effect of the ion bombardment chemical artifacts are formed on the sample surface. An enhanced catalytic reaction takes place on the continuously refreshed surface of the sample and causes carbonization of the sample surface. This effect results partly in preferential sputtering partly in a localized stress field near the sample surface. The appearance of these artifacts is very similar to those expected from direct radiation damage.

The mixing effect is illustrated in Fig. 48 which shows an example of a cross-sectionally thinned amorphous Si/Ge multilayer system. The multilayer system is composed of eight sets of 14 Si layers (2 nm thick) and 15 Ge layers (2 nm thick) separated by amorphous Si layers (10 nm thick). To avoid the near-perpendicular beam incidence at the edge of the perforation, the sample was rocked during the last period of the thinning.

The angle of incidence of the ion beam along the sample axis x is marked in Fig. 48a; it is the same on both sides of the sample. The thickness distribution of the cross-sectionally thinned sample (Fig. 48b), starting from the perforation edge, was determined using the mass-thickness method. This method was used because the structure of the sample was

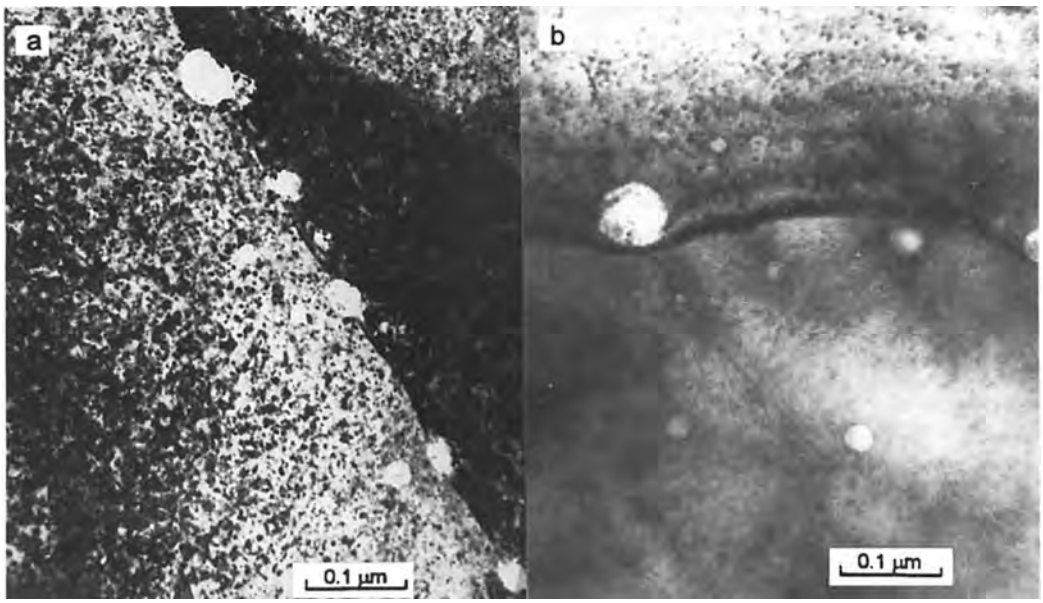


Figure 47. Ion beam thinned W samples: (a) $E = 3\,\text{keV}$, $\theta = 84°$; (b) $E = 0.5\,\text{keV}$, $\theta = 84°$.

amorphous. When the thickness of the sample is <3.5 nm the individual 2 nm broad layers are barely visible in the background due to the mixing effect. However, the 10 nm broad Si layer in the 2 nm range of the thinned sample is clearly visible. From these results one can conclude that the sphere of mixing, in both the lateral and perpendicular directions of the sample co-ordinates, cannot be larger than 1–1.5 nm.

Besides the above-mentioned phenomena (damaged-layer formation, mixing, surface catalytic reactions), some other effects should be considered for specific materials. For example, the electrical-charging effect can be a problem when insulating materials are prepared by ion milling [114], and redeposition of sputtered atoms on the eroded surface, which is a serious problem with focused ion beam (FIB) milling, should be avoided.

3.3.6 Sample Temperature during Sputtering: Application of Cooling Stages

Bombardment with high-energy ions during thinning results in a rise in the temperature of the specimen. The actual temperature depends on the sputtering conditions (energy, incidence, and intensity of the ion and neutral beam, and the atomic mass of the sputtering gas), the thermal conductivity (inside the sample, and between the sample and the sample holder), and the heat radiation (e.g. the emissivity of the sample surface).

In order to re-establish the same thinning conditions it is necessary to monitor the flux of the bombarding particles. If the beam consists only of ions it is possible to measure the current landing on the sample

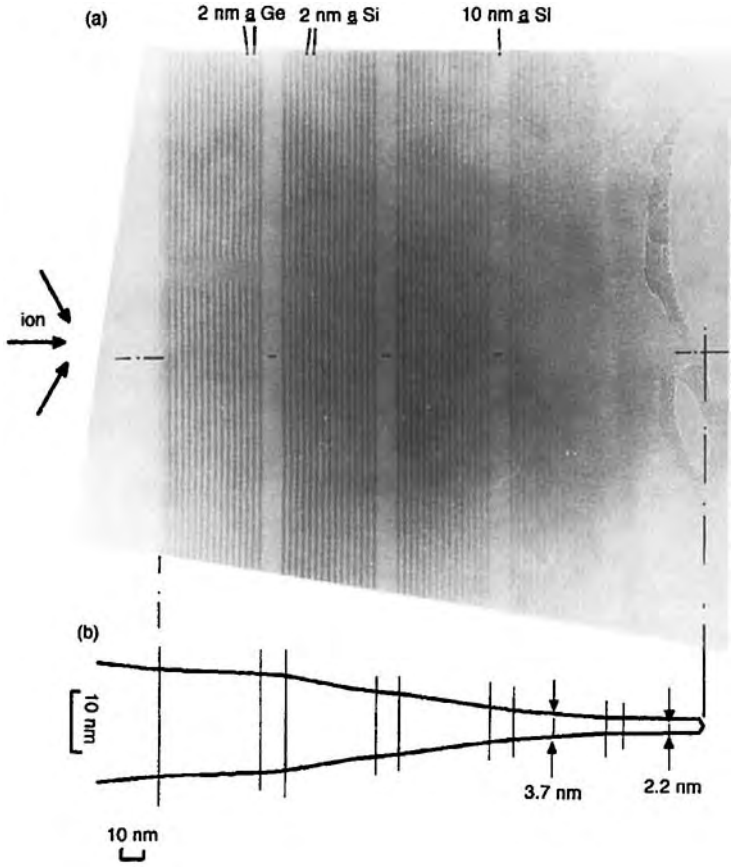


Figure 48. A TEM cross-sectional image of an amorphous Si/Ge multilayer system (a) and its measured thickness profile (b).

and the holder. However, even in this case the measurement cannot be sufficiently precise because of the unknown yield of secondary electrons. Measurement of the sample current is also useless if the beam contains predominantly neutral atoms. Therefore, in practice, the discharge power (current and anode voltage) of the ion guns is measured.

The measured data on the sample temperature during ion beam thinning [89, 115–119] are characteristic of a given type of ion milling unit and the exact application of the measured data to another unit is difficult. However, the trends of the sample temperature dependence on the thinning conditions are general:

- The sample temperature increases in rough proportion with the beam energy (anode voltage), the beam incidence angle γ [115], and the beam intensity, if the output is proportional to the discharge current of the gun [117, 119].
- The sample temperature is of the order of 100–150°C [89, 115], but much higher temperatures can easily be generated [67, 117] even in the case of metallic samples [120].

If there is a maximum temperature which can be allowed, the thinning parameters must be chosen (e.g. reduction of gun current) so that this temperature is not exceeded and so that effective thinning can take place.

It is not possible to measure the temperature of a thin sample (close to the perforation) using a thermocouple (e.g. [115, 120]) or a Thermochrom [119]. To determine the maximum temperature, the 'thin film melting method', frequently used in the field of in situ electron microscopy, can be applied. In this method a low-melting-point metal (In, Sn, Pb, etc.) or metallic alloy (e.g. AuGe) of a thickness of 10–30 nm is evaporated simultaneously onto the surface of an Si disc (for measurement) and onto a thin-film substrate (for control purposes). The Si disc is thinned from the rear side by ion beam milling and the transparent specimens are checked in the TEM. If the temperature during etching exceeds the melting point of the metal, then the appearance of the thin film on the Si disc and on the control sample are quite different.

Figure 49 shows an example of thin Pb film melting experiments done in the normal sample holder of an IV3 thinning unit (Technoorg-Linda) equipped with teletwin ion guns. When an Si disc was inserted in the sample holder (mechanical fixing, small fixing strength) and the sputtering speed was $40\,\mu m\,h^{-1}$ (two guns, 10 kV, 2 mA; $\gamma = 15°$), the Pb layer melted (temperature >328°C) (Fig. 49b). However, when the thermal conduction between the sample and the holder was increased by means of a heat-conducting metal paste, the Pb layer remained unchanged under the same thinning conditions (Fig. 49a).

When a single-sided sample holder is used, the sample can be fixed in the holder using a conducting glue, resulting in good heat conduction between the specimen and the holder [78]. In the case of low angle thinning, where the maximum sputtering speed is about $15–20\,\mu m^{-1}$ and the incident angle of the ion beam is $\gamma < 7°$, the temperature of the sample does not usually exceed 150°C. When the temperature exceeds a certain value the glue decomposes and becomes volatile in the vacuum (e.g. superglue type 'ferrobond', ~120°C; araldit type AT1, ~180°C). This phenomena can

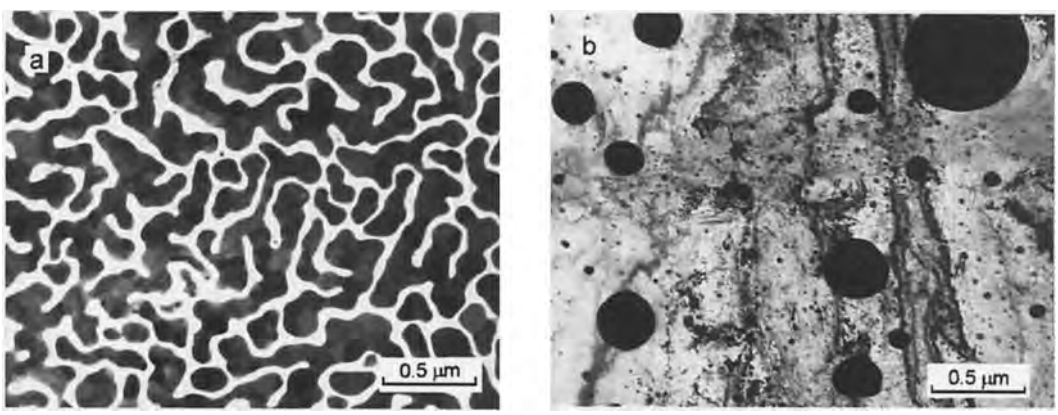


Figure 49. Images of a Pb layer when the conduction of heat between the specimen and the holder is (a) good and (b) poor.

also be a sign that the temperature limit has been exceeded during ion milling.

In the case of heat-sensitive specimens, a liquid nitrogen cooling stage can be used. The temperature of the specimen holder can be held below 100°K under the low angle thinning condition by applying an effective cooling stage, even if the guns are working at full power [121]. To avoid a significant temperature difference between the sample and the holder during thinning, the sample must be fixed with the provision of good heat flow between the sample and the sample holder. This also applies to thinned samples, where the conduction of heat between the thin part and thicker rim must be sufficient for effective cooling. It is important to reduce the power of the thinning process (anode voltage and discharge current of guns) before perforation of the specimen, in order to avoid heating of the thin area of the sample.

3.3.7 Reactive Ion Beam Thinning

This technique, also known as 'chemically assisted ion bombardment etching' [122, 123], is used when (as, typically, is the case with compound semiconductors) Ar ion milling leads to the formation of artifacts. The technique also has the advantage of reducing preferential sputtering – the incident ions react with a component of the sample being bombarded, leading to chemical reactions in addition to sputtering. Iodine is the most widely used material for reactive ion etching.

In the case of InP, the P is preferentially lost, while the excess In forms droplets on the surface of the sample [124]. In the case of II–VI compound semiconductors, much damage (dislocations, stacking faults) arises from Ar^+ ion bombardment [125, 126]. These artifacts hinder investigation of the sample structure. In some cases it is impossible to distinguish the original sample structure from the artifacts. Some improvement can be obtained by decreasing the accelerating voltage and the ion current, but these measures are not sufficient to give artifact-free samples. Liquid nitrogen cooling of the sample can decrease the number of defects, but cannot eliminate them [124].

The solution for the thinning of these sophisticated samples is to use the reactive ion milling method. The idea of reactive ion milling came from the semiconductor industry, but the application of Cl^+ ion beams and CF_4 discharges was not successful, giving an InP sample surface with accumulations of excess indium [127].

Chew and Cullis [124] have successfully applied iodine for the reactive ion milling of InP. The method produces artifact-free samples with efficient liquid nitrogen cooling [124, 128]. They proposed different processes for the preparation of InP samples, such as sputtering with a 1:1 mixture of Ar^+ and I^+, or re-milling the sample for 10 min with I^+ ions. They have also shown the beneficial effect of I^+ bombardment in the preparation of II–VI compounds [126]. Ivey and Piercy [129] have reported another solution to the problem. They used a 75% Ar/25% O_2 mixture for ion bombardment and, although the mechanism of this method is not yet clear, and it is only effective with liquid nitrogen cooling.

Yu et al. [126] have reported on the preparation of ZnS and ZnSe samples, where the application of iodine was also beneficial, but the final solution was the

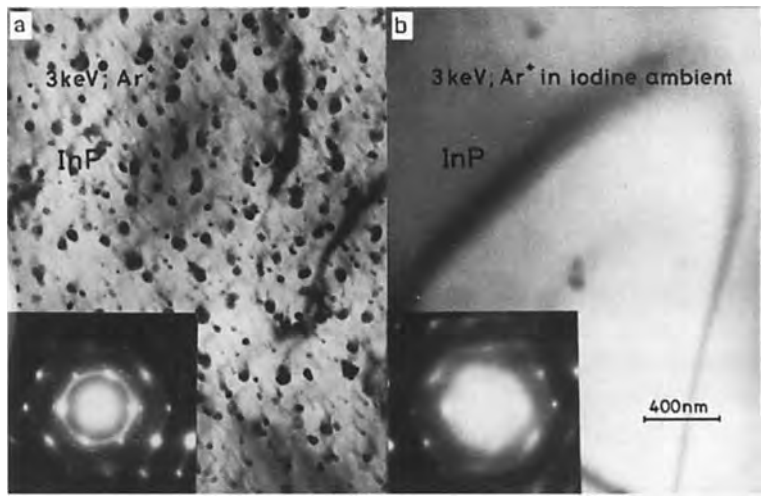


Figure 50. The surface of an InP sample: (a) bombarded with 3-keV Ar^+ ions; (b) bombarded as in (a) but in an iodine atmosphere.

chemical etching of a prethinned specimen in a Br_2/CH_3OH solution.

Reactive ion etching can be achieved without ionized reactive material, simply by introducing the reactive gas into the vacuum chamber of the ion beam milling unit [130, 131]. For example, a sublimating iodine crystal can be used for this purpose, with a precise valve being used to control the partial pressure of the reactive gas.

The surface of an InP sample thinned by Ar^+ ions is shown in Fig. 50a. Despite the fact that the sample was bombarded with the beam at a very low angle of incidence during the whole thinning process and with 3-keV Ar^+ ions only in the final stage (10–15 min), the resulting surface structure is very rough and excess In droplets can be observed on the surface of the sample. The artifact-free InP sample shown in Fig. 50b was bombarded with Ar^+ ions at 10 keV until perforation of the sample. The accelerating voltage was then decreased to 3 keV and iodine was let in to the vacuum chamber at a partial pressure of 1×10^{-4} mbar.

A cross-sectional high resolution electron microscopy (HREM) image of an InP/InGaAs sample prepared by Ar ion milling in an iodine atmosphere is shown in Fig. 51. It can be seen that the surface is artifact free and the interfaces can be studied at high resolution.

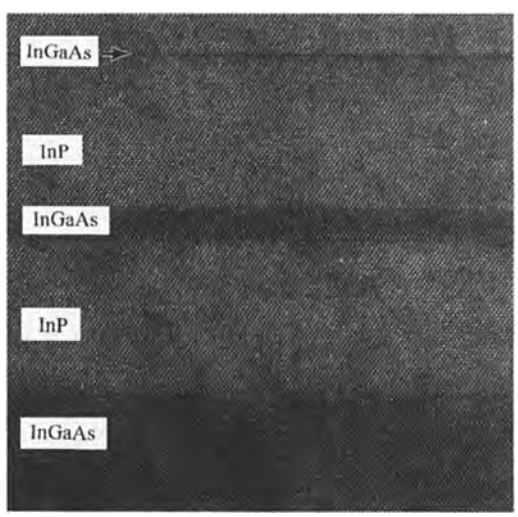


Figure 51. Cross-sectional image of an InP/InGaAs quantum-well structure after treatment with iodine. (Courtesy of Dr Jan-Olle Malm, National Center for HREM, Lund University, Sweden.)

The explanation of the reactive ion etching of InP in an iodine atmosphere is as follows. The iodine reacts with In to form compound, which has a higher sputtering yield than In. This reaction product can be removed very easily with 3-keV Ar^+ ions. During the reactive ion milling process a thin (10–50 nm) layer must be removed from the sample. It is not necessary to apply low angle ion beam milling [58] or sample cooling [130, 131].

Low ion energies and small angles of incidence of the beam both decrease the ion range in the solid, which leads to a minimization of artifact formation.

All important compound semiconductors can be thinned using ion milling, except ZnS and ZnSe for which Ar^+ ion beam milling should be followed by chemical etching for a few seconds [132].

Oxygen assisted ion thinning of diamond/Si samples has been realized and described [133]. The large difference in the sputtering rates of Si and diamond was overcome by Ar^+ milling at a low angle of incidence (see Sec. 3.3.4). However, Ar^+ milling in an O_2 atmosphere increased the milling rate of diamonds about three-fold, leaving the milling rate of Si practically unchanged. The explanation for this effect is the formation of volatile compounds (CO, CO_2) under ion bombardment [133].

3.3.8 Ion Beam Slope Cutting (Cross-Sectional Sample Preparation) for Scanning Electron Microscopy

This method was developed to produce microcuts through heterogeneous materials [134, 135] in a specific area, particularly for use in SEM investigations. This method has essential advantages in comparison with mechanical grinding methods and can be applied to most substances (e.g. materials composed of very different components, porous and brittle materials, etc.). If the ion bombardment is carried out in situ in a scanning electron microscope, the cutting can be controlled and the cut plane observed. A survey of this method has been given by Hauffe [135]. The ion beam slope cutting method can also be used to cut a thin foil from a specific area of a sample to obtain an XTEM specimen [136].

The method consists of applying a collimated ion beam with good directional stability, with a part of the sample screened by a smooth edge along the desired line of cutting. The teletwin ion guns fulfil the requirements for this technique, and therefore the method can also be applied in a thinning unit, using a slope cutting sample holder. A sketch of such a holder is shown in Fig. 52. The cut facet is inclined at 45° to the sample surface. This geometry is generally suitable for SEM investigation. To avoid a furrowed surface structure of the cut facet, the sample is rocked during ion beam cutting. The cutting process takes about 30 min.

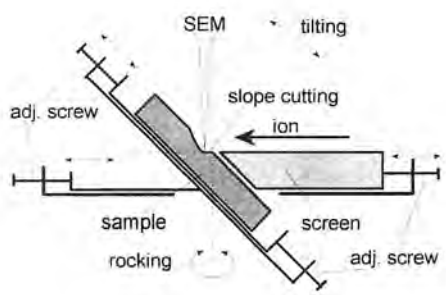


Figure 52. A sample holder for slope cutting for SEM.

The advantages of the method are:

- The surface to be investigated is free of mechanical damage.
- There is practically no radiation damage, because the ions strike the surface at glancing incidence ($\gamma \leqslant 1.5°$).
- The surface to be investigated is also smooth in the case of heterogeneous samples, because of the high step movement speed on the cut surface. A short period of selective etching can be applied to reveal the internal structure by tilting the cut facet perpendicular to the ion beam.
- High accuracy ($\leqslant 5\,\mu m$) positioning of the cut is possible.

An example of the slope cutting method for p–n junction localization in integrated circuits [137] is shown in Fig. 53.

3.3.9 Focused Ion Beam Sample Preparation for Transmission Electron Microscopy

The TEM characterization of localized, submicrometer areas is a demand that has arisen in the microelectronics field. A focused ion beam (FIB) milling provides a means for this kind of sample preparation.

High energy ion sources are used in FIB systems. Most frequently Ga liquid metal ion sources operating at 25–35 keV are applied. The units have several fields of application, such as imaging, device modification, sample preparation for SEM and TEM, and sometimes SIMS. A comprehensive survey of these options is given by Stevie et al. [138]. Images of the sample can be obtained by the detection of secondary electrons while the surface is scanned by

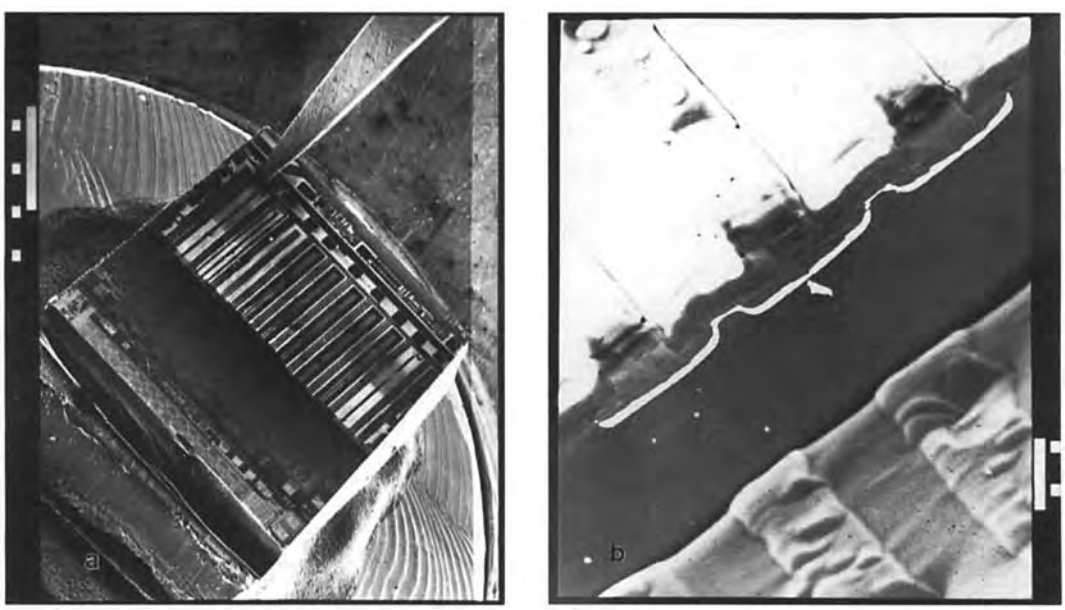


Figure 53. SEM images of an integrated circuit sample obtained by slope cutting: (a) low magnification; (b) hgher magnification image with electron beam induced current signal.

the ion beam. Device modification means that material can be removed or deposited selectively onto the sample, and electronic devices (masks and circuits) can be repaired in this way (this topic is outside the scope of this Chapter).

FIB can be used to prepare samples for SEM and TEM, because where the surface is hit by the scanned beam some material is removed from the surface. By changing the beam current, sputtering rates of 0.005–1 $\mu m^3 s^{-1}$ can be obtained; the sputtering rate also depends on the type of material to be removed. The erosion speed is a function of the area being scanned. The resolution of this technique is 50 nm, which is excellent for the location and preparation of submicrometer details of a sample. When the region of interest has been located, the sample is bombarded perpendicular to the surface, moving closer and closer to the region of interest, from both sides. Usually the sputtering process can be controlled by means of a computer program. Using this technique a substantial amount of material needs to be removed to cut out a membrane for TEM investigations. The total sputtering time has been decreased by modifying the method such that staircase-shaped recessions are sputtered on both sides of the region of interest [139]. Mesa etching is also used to locate the area of interest and to remove some material [140]. Ben Assayag et al. [140] cut mesa etched structures using a diamond saw and mounted them on a TEM grid before FIB bombardment. This technique allows easy handling of the thinned specimens. When the region of interest (the TEM sample) is sputtered directly from the whole wafer, redeposition of the sputtered atoms occurs during the FIB process because the side walls of the cut region are very close to the sample surface. The problem of redeposition was avoided in the above-mentioned case [140], because the region of interest was cut away and the other remainder of the wafer removed before the FIB process. A simple solution to the problem has been proposed by Tarutani et al. [141]. By cleaving off a small triangular-shaped piece of sample, the redeposition of sputtered atoms could be avoided. Yamaguchi et al. [142] describe an elegant method of cutting a thin slice (20 μm) of a specimen for FIB by using a diamond saw.

In many cases a metallic covering layer is deposited on the region of interest [139–141, 143] to protect the top part of the structure during bombardment. Usually W is used for this purpose, being applied by means of the selective deposition option of FIB. The metal layer also prevents the charging effect of ion bombardment.

The sample is usually tapered during bombardment, but this effect can be eliminated [142] by tilting the FIB sample holder.

It is advantageous to mount the sample on a TEM grid before FIB sputtering [140], although sensitive techniques are available for removing the thinned membrane sample after sputtering [139].

FIB allows for the preparation of TEM specimens comprising selected, local, submicrometer regions, which are difficult to obtain by means of the conventional ion beam milling techniques.

3.3.10 Ion Beam Polishing for Scanning Tunneling Microscopy

Scanning tunneling microscopy (STM) and atomic force microscopy (AFM) have

recently become widely used for surface characterization. Using these methods well-defined surfaces can be characterized on an atomic scale. When the task is to study the cross-section of a layered structure, sample preparation is crucial. Ion beam milling of embedded samples can provide samples that can be used for scanning probe microscopy (SPM), although atomic resolution cannot be expected due to the damaged surface reconstruction.

Ion beam etching has been used to prepare cross-sections for SPM from hard coatings [144, 145]. The method for preparing samples for TEM has been modified as only one of the surfaces should be polished for SPM. The sample to be investigated is placed face to face with a slice of Si and the pieces embedded together. After mechanical grinding the specimens are bombarded for 1 h with 10-kV Ar^+ ions at an angle of incidence of 3°. Sample rocking is applied from the silicon side (with one gun) to translate the smooth silicon surface to the hard coating (e.g. to remove the surface steps from the area of interest). In the final 10 min the accelerating voltage of the ion gun is decreased to 2.5 kV and the ion current to 1 mA (from 3 mA). To develop the surface topography (grain boundaries) additional ion beam etching at a higher angle of incidence of the beam (25°) can be used at 2 kV for 2.5 min.

Samples prepared in this way can be used for AFM, although applying STM to localize the appropriate region has caused problems. TiN hard coatings and Mo/Ti layers have been investigated in cross-section [144, 145]; the grain structure could be resolved easily.

The SPM of samples prepared using the above-described method to solve problems of material characterization can provide useful information, although the atomic resolution obtained with surface characterization techniques (STM and AFM) cannot be achieved in this way.

3.4 References

[1] P. J. Goodhew in *Practical Methods in Electron Microscopy*, Vol. 1 (Ed.: A. M. Glauert), North-Holland, London **1972**, p. 1.

[2] H. M. Wilson, A. J. Rowe in *Practical Methods in Electron Microscopy*, Vol. 8 (Ed.: A. M. Glauert), North-Holland, London **1980**.

[3] S. R. Keown in *Modern Physical Techniques in Materials Technology* (Ed.: T. Mulvay, R. K. Webster), Oxford University Press, Oxford **1974**, p. 94.

[4] D. E. Bradley, *Nature* **1959**, *181*, 875.

[5] K. K. Adachi, M. Hojou, M. Katoh, K. Kanaya, *Ultramicroscopy* **1976**, *2*, 17.

[6] H. Abbink, R. M. Broudy, G. P. McCarthy, *J. Appl. Phys.* **1968**, *39*, 4673.

[7] J. Kistler, U. Abei, E. Kellenberger, *J. Ultrastruct. Res.* **1977**, *59*, 76.

[8] A. W. Agar, R. H. Andreson, D. Chescoe in *Practical Methods in Electron Microscopy*, Vol. 2 (Ed.: A. M. Glauert), North-Holland, London **1974**.

[9] F. Keller, A. H. Giessler, *J. Appl. Phys.* **1944**, *34*, 345.

[10] N. C. Welsh, *J. Inst. Metals* **1956**, *85*, 129.

[11] American Society for Testing and Materials, *ASTM Metals Handbook*, Vol. 9, 8th edn., ASTM **1974**, 58.

[12] L. Young, *Anodic Oxide Films*, Academic Press, New York, **1961**.

[13] B. A. Smith, J. Nuttig, *Br. J. Appl. Phys.* **1956**, *7*, 427.

[14] A. Cyrska-Filemonowicz, K. Spiradek, *Eur. Microsc. Anal.* **1995**, *3*, 13.

[15] G. J. C. Carpenter, J. NG-Yelim, M. W. Phaneuf, *Microsc. Res. Technique* **1994**, *28*, 422.

[16] T. Habrovec, *Prakt. Metallogr.* **1968**, *5*, 18.

[17] G. H. Karcher, E. T. Stephenson, *Trans. Am. Soc. Metals* **1967**, *60*, 716.

[18] G. A. Basset, *Phil. Mag.* **1958**, *3*, 1042.

[19] L. A. Shenyavskaya, G. I. Distler, *Fiz. Tverd. Tela* **1976**, *18*, 1451.

[20] U. Essman, H. Trauble, *Phys. Lett. A* **1967**, *24*, 526.

[21] H. Bethge, *Phys. Stat. Sol.* **1962,** *2*, 3.
[22] H. Bethge, *Phys. Stat. Sol.* **1962,** *2*, 775.
[23] W. M. Kosevich, A. A. Sokol, G. Radnóczi, *Acta Phys. Hung.* **1973,** *33*, 353.
[24] S. Amelinckx, P. Delavignette in *Direct Observations of Imperfections in Crystals* (Eds.: J. B. Newkirk, J. H. Wernick), Interscience, New York **1962,** p. 295.
[25] P. B. Price, R. M. Walker, *J. Appl. Phys.* **1962,** *33*, 3400.
[26] G. W. Farrants, A. K. Reith, B. Schüller, K. Feren, *J. Microsc.* **1988,** *149*, 159.
[27] M. Fendorf, M. Powers, R. Gronsky, *Microsc. Res. Technique* **1995,** *30*, 167.
[28] A. Malchére, E. Gaffet, *Proc. 13th International Congress on Electron Microscopy*, Vol. 1 (Eds.: B. Jouffrey, C. Coliex), Editions de Physique, Les Ulis **1994,** 1025.
[29] G. Vincze, Á. Barna, G. Radnóczi, *Proc. Multinational Congress on Electron Microscopy*, Parma, Italy **1993,** 349.
[30] G. Vincze, A. Barna, G. Radnóczi, *Proc. 13th Int. Conf. on Electron Microscopy*, Paris **1994,** 1001.
[31] A. Pintér Csordás-Tóth, Á. Csanády, G. Vincze, *Proc. Multinational Congress on Electron Microscopy*, Parma, Italy **1993,** 297.
[32] Y. O. Kim, A. Ourmazd, *Proc. 13th International Congress on Electron Microscopy*, Vol. 1 (Eds.: B. Jouffrey, C. Coliex), Editions de Physique, Les Ulis **1994,** 1017.
[33] A. Ourmazd, *Mater. Sci. Rep.* **1993,** *9*, 201.
[34] Y. Kim, A. Ourmazd, M. Bode, R. D. Feldman, *Phys. Rep. Lett.* **1989,** *63*, 636.
[35] I. A. Bucklow, M. Cole, *Met. Rev.* **1969,** *14*, 103.
[36] P. B. Hirsch, R. B. Nicholson, A. Howie, D. W. Pashley, M. J. Whelan, *Electron Microscopy of Thin Crystals*, London, Butterworths **1965,** pp. 43, 50–63.
[37] H. M. Tomlinson, *Phil. Mag.* **1958,** *3*, 867.
[38] P. B. Hirsch, R. B. Nicholson, A. Howie, D. W. Pashley, M. J. Whelan, *Electron Microscopy of Thin Crystals*, London, Butterworths **1965,** Appendix 1, p. 453, and references therein.
[39] K. C. Thompson-Russel, J. W. Edington, *Practical Electron Microscopy in Material Science, Monograph 5: Electron Microscopy Preparation Techniques in Materials Science*, N. V. Philips' Gloeilampenfabriken, Eindhoven **1977.**
[40] P. R. Strutt, *Rev. Sci. Instrum.* **1961,** *32*, 411.
[41] W. Bollman, *Phys. Rev.* **1956,** *103*, 1588.
[42] P. J. Goodhew in *Practical Methods in Electron Microscopy*, Vol. 1 (Ed.: A. M. Glauert), North-Holland, London **1972,** pp. 98–110, and references therein.
[43] P. J. Goodhew in *Practical Methods in Electron Microscopy*, Vol. 11 (Ed.: A. M. Glauert), North-Holland, London **1985.**
[44] W. M. Rainforth, A. R. Thölén, *Procedures in Electron Microscopy* (Eds.: A. W. Robards, A. J. Wilson), Wiley, Chichester **1993,** para. 6.2.
[45] P. J. Goodhew in *Practical Methods in Electron Microscopy*, Vol. 1 (Ed.: A. M. Glauert), North-Holland, London **1972,** p. 50.
[46] Á. Barna, G. Radnóczi, *Zavodsk. Lab.* **1980,** *46*, 741.
[47] T. F. Malis, D. Steele in *Workshop on Specimen Preparation for TEM of Materials II. MRS Symposium Proceedings*, Vol. 199 (Ed.: R. Anderson), MRS, Pittsburgh **1990.**
[48] H. J. Fecht, E. Hellstern, Z. Fu, W. L. Johnson, *Met. Trans. A* **1990,** *21*, 2333.
[49] S. R. Glanvill, *Procedures in Electron Microscopy* (Eds.: A. W. Robards, A. J. Wilson), Wiley, Chichester **1993,** para. 6.6.
[50] A. M. Glauert, R. Philips, in *Techniques for Electron Microscopy* (Ed.: D. A. Kay), Blackwells, Oxford **1965, p.** 213.
[51] J. P. Heuer, D. G. Howitt, *J. Electron Microsc. Technique* **1990,** *14*, 79.
[52] R. C. Furneaux, G. E. Thompson, G. C. Wood, *Corrosion Sci.* **1978,** *18*, 853.
[53] H. Kakibayashi, F. Nagata, *Jpn. J. Appl. Phys.* **1985,** *24*, L905.
[54] P. Rutera, J. D. Ganiére, P. A. Buffat, *J. Microsc. Spectrosc. Electron* **1988,** *13*, 421.
[55] J. P. McCaffery, *Ultramicroscopy* **1991,** *38*, 149.
[56] J. P. McCaffery, *Microsc. Res. Technique* **1993,** *24*, 180.
[57] J. P. McCaffery, *Mat. Res. Soc. Proc.* **1992,** *254*, 109.
[58] Á. Barna, *Mat. Res. Soc. Symp. Proc. Ser.* **1992,** *254*, 3.
[59] R. Cong-Xin, C. Guo-Ming, F. Xin-Ding, Y. Jie, F. Hong-Li and T. Shih-Chang, *Radiation Effects* **1983,** *77*, 177.
[60] C. W. Bulle-Lieuwma, P. C. Zalm, *Surf. Interface Anal.* **1987,** *10*, 210.
[61] G. K. Wehner, *J. Vac. Sci. Technol. A* **1985,** *3*, 1821.
[62] J. A. Kubby, B. M. Siegel, *Nucl. Instrum. Methods Phys. Res. B* **1986,** *13*, 319.
[63] H. J. Roosendaal in *Sputtering by Particle Bombardment*, Vol. I (Ed.: R. Becrisch), Springer, Berlin **1981.**
[64] M. A. Tagg, R. Smith, J. M. Walls, *J. Mater. Sci.* **1986,** *21*, 123.
[65] H. H. Andersen, H. L. Bay in *Sputtering by Particle Bombardment*, Vol. I (Ed.: R. Becrisch), Springer, Berlin **1981.**

[66] P. J. Goodhew in *Practical Methods in Electron Microscopy* (Ed.: A. M. Glauert), Elsevier, Amsterdam **1985**.
[67] L. Malicskó, Á. Barna, R. Scholz, *Cryst. Res. Technol.* **1988,** *23*, 223.
[68] Á. Barna, P. B. Barna, A. Zalar, *Vacuum* **1990,** *40*, 115.
[69] Á. Barna, P. B. Barna, A. Zalar, *Surf. Interface Anal.* **1988,** *12*, 144.
[70] Á. Barna in *Proc. Eighth European Congress on Electron Microscopy*, Budapest, Hungary, Vol. 1 (Eds.: Á. Csanády, P. Röhlich, P. Szabó) **1984,** p. 107.
[71] J. P. Ducommun, M. Cantagrel, M. Moulin, *J. Mater. Sci.* **1975,** *10*, 52.
[72] I. H. Wilson in *Proc. Invited Lectures, 7th Yugoslav Symposium and Summer School on The Physics of Ionised Gases*, Rovinj (Ed.: V. Vunjovic), Institute of Physics, Zagreb **1974,** p. 589.
[73] O. Auciello, *J. Vac. Sci. Technol.* **1981,** *19*, 841.
[74] Y. Yokota, H. Hashimoto, M. Song, M. Awaji, *Jpn. J. Appl. Phys.* **1990,** *29*, 739.
[75] J. Y. Yao, G. L. Dunlop, *J. Electron Microsc. Technique* **1991,** *19*, 90.
[76] U. Helmersson, J. E. Sundgren, *J. Electron Microsc. Technique* **1986,** *4*, 361.
[77] F. Shaapur, K. A. Watson, *Mat. Res. Soc. Symp. Proc.* **1992,** *254*, 153.
[78] Á. Barna, L. Gosztola, Gy. Reisinger, *Hungarian Patent* No. 205814, 1989.
[79] J. Vanhellemont, H. Bender, C. Claeys, J. Van Landuyt, G. Declerc, S. Amelinckx, R. Van Overstraeten, *Ultramicroscopy* **1983,** *11*, 303.
[80] J. C. Brawman, R. Sinclair, *J. Electron Microsc. Technique* **1984,** *1*, 53.
[81] S. J. Klepeis, J. Benedict, R. M. Anderson, *Mat. Res. Symp. Proc.* **1988,** *115*, 179.
[82] J. Benedict, R. M. Anderson, S. J. Klepeis, *Mat. Res. Symp. Proc.* **1992,** *254*, 121.
[83] J. Benedict, S. J. Klepeis, W. G. Vandigrift, R. M. Anderson in *Technical Manual*, Tripod Polisher, South Bay Technology, San Clemente, USA.
[84] R. A. Alani, P. R. Swann, *Mat. Res. Symp. Proc.* **1992,** *254*, 43.
[85] V. L. Carlino, A. J. Hidalgo, *Mat. Res. Symp. Proc.* **1992,** *254*, 99.
[86] P. E. Fischione, E. A. Fischione, T. F. Kelly, A. M. Dalley, L. M. Holzman, D. Dawson-Elli, *Mat. Res. Symp. Proc.* **1992,** *254*, 79.
[87] A. Romano, J. Vanhelemot, H. Bender, J. R. Morante, *Ultramicroscopy* **1989,** *31*, 183.
[88] M. C. Maden, P. Crafard, *J. Electron Microsc. Technique* **1989,** *11*, 161.
[89] D. J. Barber, *J. Mat. Sci.* **1970,** *5*, 1.
[90] C. G. Crockett, *Vacuum* **1972,** *3*, 11.
[91] H. Bach, *BOSH Techn. Ber.* **1964,** *1*, 1.
[92] H. Bach in *Methodensamlung der elektronenmikroskopie* (Eds.: G. Shimmel, W. Vogel), Wissenschaftliche, Stuttgart **1970–1973,** p. 2.4.2.1.
[93] K. Yoshida, T. Yamada, *Rev. Sci. Instrum.* **1983,** *55*, 551.
[94] J. Francs, A. M. Ghander, *Vacuum* **1974,** *24*, 489.
[95] J. Francs, *J. Vac. Sci. Technol.* **1979,** *16*, 181.
[96] Á. Barna, Gy Reisinger, L. Zsoldos, *Hungarian Patent* No. 190855 **1983**.
[97] J. Francs, *Vacuum* **1984,** *34*, 259.
[98] R. Messier, K. E. Spear, A. R. Badzian, R. Roy, *J. Metals* **1987,** *39*, 8.
[99] L. Tóth, personal communication.
[100] K. Kanaya, K. Honnjou, K. Adachi, K. Toki, *Micron* **1974,** *5*, 89.
[101] D. J. Clinton, *Micron* **1972,** *3*, 358.
[102] S. J. Zincle, C. P. Haltom, L. C. Jenkins, C. K. H. DuBose, *J. Electron Microsc. Technique* **1991,** *19*, 452.
[103] O. Unal, A. H. Heuer, T. E. Mitchel, *J. Electron Microsc. Technique* **1990,** *14*, 307.
[104] H. L. Humiston, B. M. Tracy, M. Lawrance, A. Dass, *Mater. Res. Symp. Proc.* **1992,** *254*, 221.
[105] K. Ostreicher, C. Sung, *Mater. Res. Symp. Proc.* **1992,** *254*, 193.
[106] C. Wan, M. Dupeux, *Microsc. Res. Technique* **1992,** *23*, 228.
[107] G. Sáfrán, P. Panine, *Microsc. Res. Technique* **1993,** *25*, 346.
[108] D. J. Barber, *Ultramicroscopy* **1993,** *52*, 101.
[109] T. Diaz de la Rubia, M. W. Gunian, *Phys. Rev. Lett.* **1991,** *66*, 2766.
[110] G. Van Tendeloo, M.-O. Ruault, H. Bernas, M. Gasgnier, *J. Mater. Res.* **1991,** *6*, 677.
[111] G. Carter, J. Colligon, *The Ion Bombardment of Solids*, Elsevier, Amsterdam **1968**.
[112] T. Ishiguro, T. Susuki, N. Susuki, M. Ozawa, *J. Electron Microsc.* **1987,** *36*, 163.
[113] P. C. Zalm, C. J. Vriezema, *Nucl. Instrum. Methods Phys. Res. B* **1992,** *67*, 495.
[114] T. R. Lundquist, R. Alani, P. R. Swann, *Ultramicroscopy* **1988,** *24*, 27.
[115] D. Bahnck, R. Hull, *MRS Symp. Proc.* **1990,** *199*, 253.
[116] G. Lu, F. Niu, R. Wang, *Rad. Effects* **1991,** *116*, 81.
[117] G. Lu in *Proc. of Asia-Pacific EM Conf. Beijing*, China (Eds.: K. H. Kuo, Z. H. Zhai), **1992,** *1*, 556.
[118] G. Lu, *Phil. Mag. Lett.* **1993,** *68*, 1.
[119] D. Bahnck, R. Hull, *MRS Symp. Proc.* **1992,** *254*, 249.
[120] M. J. Kim, R. W. Carpenter, *Ultramicroscopy* **1978,** *21*, 327.

[121] Á. Barna in *Workshop on TEM Specimen Preparation by Ion and Atom Beam Milling*, Budapest **1992,** unpublished.
[122] J. Roth in *Sputtering by Particle Bombardment*, Vol. 2 (Ed.: R. Berish), Springer, Berlin **1983,** p. 91.
[123] S. P. Withrow, D. B. Poker (Eds.), *Proc. 7th Int. Conf. on Ion Beam Modification of Materials*, Knoxville, USA, 9–14 September 1990, *Nucl. Instrum. Methods B* **1991,** *59/60*, 1.
[124] N. G. Chew, A. G. Cullis, *Ultramicroscopy* **1987,** *23*, 175.
[125] A. G. Cullis, N. G. Chew, J. L. Hutchinson, *Ultramicroscopy* **1985,** *17*, 203.
[126] J. E. Yu, K. S. Jones, R. M. Park, *J. Electron Microsc. Technique* **1991,** *18*, 315.
[127] R. A. Baker, T. M. Mayer, R. H. Burton, *Appl. Phys. Lett.* **1982,** *40*, 583.
[128] N. G. Chew, A. G. Cullis, *Appl. Phys. Lett.* **1984,** *44*, 142.
[129] D. G. Ivey, G. R. Piercy, *J. Electron Microsc. Technique* **1987,** *5*, 107.
[130] Á. Barna, B. Pécz, *J. Electron Microsc. Technique* **1991,** *18*, 325.
[131] B. Pécz, Á. Barna, *Vacuum* **1993,** *45*, 1.
[132] J. E. Yu, K. S. Jones, R. M. Park, *J. Electron Microsc. Technique* **1991,** *18*, 315.
[133] P. Würzinger, M. Joksh, P. Pongratz in *Proc. of ICEM 13* (Ed. by B. Jouffrey and C. Colliex), Les Editions de Physique Les Ulis, Paris, 17–22 July **1994,** *1*, 1015.
[134] W. Hauffe, Thesis B, Dresden, Technical University **1978.**
[135] W. Hauffe in *Sputtering by Particle Bombardment*, Vol. III (Ed.: R. Berish), Springer, Berlin **1991.**
[136] W. Hauffe, *Proc. Phys. Halbleiter Oberfläche* **1989,** *20*, 229.
[137] K. Mádl, A. L. Tóth, Á. Barna, *Inst. Phys. Conf. Ser.* **1988,** *93*, 65.
[138] F. A. Stevie, T. C. Shane, P. M. Kahora, R. Hull, D. Bahnck, V. C. Kannan, E. David, *Surf. Interface Anal.* **1995,** *23*, 61.
[139] M. H. F. Overwijk, F. C. van den Heuvel, C. W. T. Bulle-Lieuwma, *J. Vac. Sci. Technol. B* **1993,** *11*, 2021.
[140] G. Ben Assayag, C. Vieu, J. Gierak, H. Chaabane, A. Pepin, P. Henoc, *J. Vac. Sci. Technol. B* **1993,** *11*, 531.
[141] M. Tarutani, Y. Takai, R. Shimizu, *Jpn. J. Appl. Phys.* **1992,** *31*, L1305.
[142] A. Yamaguchi, M. Shibata, T. Hashinaga, *J. Vac. Sci. Technol. B* **1993,** *11*, 2016.
[143] D. P. Basile, R. Boylan, B. Baker, K. Hayes, D. Soza, *Mat. Res. Soc. Symp. Proc. Ser.* **1992,** *254*, 23.
[144] K. A. Pishow, M. Adamik, P. B. Barna, A. S. Korhonen, *J. Vac. Sci. Technol. B* **1994,** *12*, 1716.
[145] K. A. Pishow, A. S. Korhonen, M. Adamik, P. B. Barna, *Surf. Coatings Technol.* **1994,** *67*, 95.

Handbook of Microscopy
S. Amelinckx, D. van Dyck,J. van Landuyt , G. van Tendeloo

4 Environmental Problems

4.1 Introduction

Although microbeam techniques are very useful for the analysis of individual environmental particles, they are still seldom used. This is surprising and may be related to the high cost of some of the instruments. Yet, the importance of particulate matter in the environment is well known: many components of environmental particles are toxic, particles are the major material carriers in water and air, atmospheric aerosols influence the visibility and the global climate, etc. Individual particle analysis is a very useful complement to conventional bulk analysis, because it can provide detailed information on the origin, formation processes, transport, reactivity, transformation reactions, and environmental impact of particles.

However, specific problems arise which are often related to the small size of the samples. Quantitative analyses are difficult to perform because the small dimension of the particles make the determination of the interaction volume uncertain and because of the absence of suitable standards. Statistically relevant results can only be obtained by the analysis of large numbers of particles per sample, which makes this very time consuming. This problem has already partly been solved by the recent developments in automation and computerization, and computer-controlled electron probe X-ray microanalysis (EPXMA) is the most advanced example of this technology. However, the major drawback of most microbeam techniques is the necessity to operate in vacuum, which implicates loss or transformation of most volatile and unstable compounds. Fourier transform infrared microscopy (FT-IR) and micro-Raman spectrometry both operate under atmospheric pressure, and are thus the only ones suitable for the analysis of these particles. Moreover, all microbeam techniques have specific limitations and they mostly complement each other with respect to detectable elements, detection limits, lateral resolution, etc. Hence, an optimal characterization of a sample is only possible by combining different microbeam techniques with bulk analysis.

This Chapter gives an overview of the recent applications of microbeam techniques, discussed in Vol. I of this book, on individual environmental particles. Environmental applications of microanalytical techniques have been reviewed by Grasserbauer [1], Van Grieken et al. [2], Van Grieken and Xhoffer [3], Xhoffer et al. [4], and Jambers et al. [5].

4.2 Scanning Microanalysis

As the differences between EPXMA and scanning electron microscopy/energy dispersive X-ray detection (SEM-EDX) have been minimized over the years to a slightly different instrumental set-up, both techniques can to some extent be used for both chemical and morphological studies. The applications of both are discussed together.

4.2.1 Automation

By combining an EPXMA or SEM-EDX system with a system for automated particle recognition and characterization (PRC), huge amounts of individual particles can be analyzed in a few hours. At the Micro and Trace Analysis Center (MiTAC) of the University of Antwerp, this automation has been achieved using home-made software. The particles are located by performing successive horizontal line scans over the selected sample area. During this scan, the contour points of all detected particles are saved. When all contour points of a certain particle are found, the area, perimeter, and equivalent diameter (i.e. the diameter of a circle which has the same area as the particle) are calculated. According to the preselected parameters, the X-ray spectrum is recorded at the center of the particle or during a star scan over the particle. After this location, sizing, and characterization, the horizontal line scans are resumed until the next particle is detected. Since this technique produces a huge data matrix, combination with multivariate techniques and/or cluster analysis is necessary.

4.2.2 Applications to Aerosols

EPXMA and SEM-EDX have been used successfully for the analysis of aerosols collected in sampling places ranging from remote to workplace environments.

The most remote place on Earth is probably the Antarctic continent, and it is thus the ideal site at which to collect background aerosols. Early research on single Antarctic aerosols from different locations have shown that the background aerosols mainly exist of sulfur rich particles (probably formed by gas-to-particle conversion), sea salt particles (formed by bubble bursting and wave action in the ocean), aluminosilicates (Earth crust dust and/or particles from local sources, such as volcanos and geysers), and iron rich particles (long-range transport anthropogenic particles and/or meteoric dust) [6–11]. The relative concentrations of these particle types vary with season, meteorological conditions, and sampling site. Aerosols collected on the coast of the Antarctic continent are dominated by marine components in both the fine and coarse fractions [12–15]. In a significant number of sea salt particles, small amounts of sulfur were detected, indicating a possible reaction with gaseous sulfur compounds. In most particles larger than 0.1 μm, sulfur is present as gypsum. Manual analysis indicated that some aerosols are an internal mixture of marine components and silicates. An example of the dependence of the season is the dominance, by number and mass, of sulfate particles during the summer [8–14, 17]. Some researchers [9, 16] have also detected H_2SO_4 and $(NH_4)_2SO_4$ particles.

By combining SEM analysis with chemical tests for the characterization of

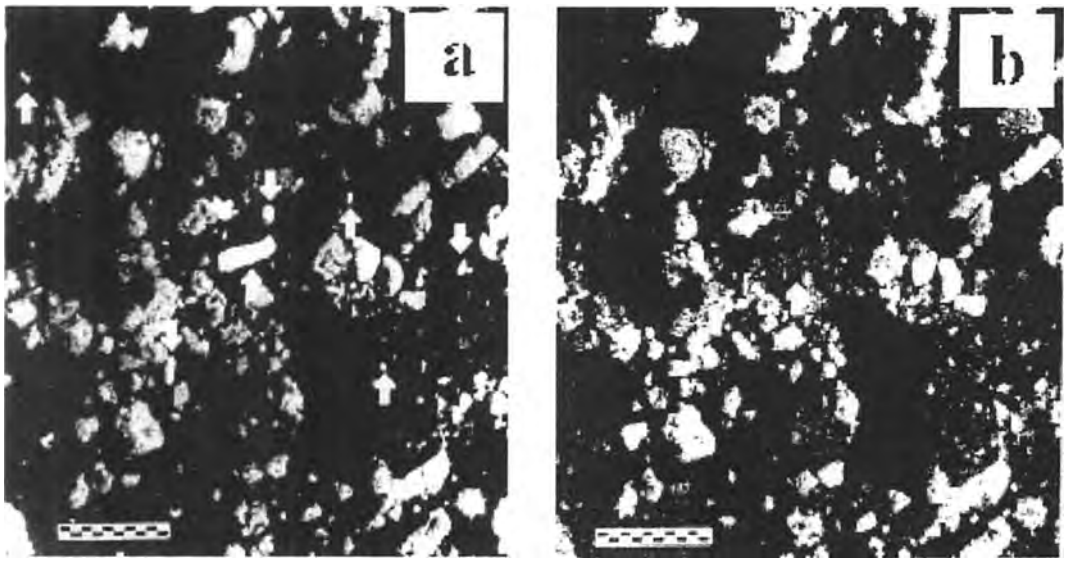


Figure 1. (a) Silicon (SiK_α) and (b) iron (FeK_α) X-ray mappings of a grain mount of silt particles from Pu'u o Mahana, Hawaii [19]. Reproduced by permission of Elsevier Science Publishers B.V.

individual aerosols from Cape Grim (Tasmania), the Mauna Loa Observatory (Hawaii) and Point Barrow (Alaska), Bigg [16] found that the majority of the particles were composed of sulfuric acid or ammonium sulfate. The sulfuric acid particles dominated the samples from Barrow and Mauna Loa, while at the Antarctic and Cape Grime sites ammonium sulfate was dominant.

Siberia is characterized by both very remote areas and some strongly polluted regions. Since aerosols from this huge subcontinent have an impact on the Arctic and on the global climate, studies have been started to investigate the production, transport, and deposition of both natural and pollution aerosols at the local, regional and long-range levels [18].

Analysis of volcanic loess on Hawaii has given new evidence of the long-range transportation of particles [19]. In late Quaternary loess deposits of the Pahala Formation at Pu'u Mahana, quartz grains were found which are virtually unknown in the basalt of Hawaii. The study was done by combining X-ray diffraction of bulk samples with SEM-EDX measurements on single mineral grains and X-ray mappings of particles in a grain matrix. Figure 1 shows a silicon and an iron X-ray mapping of the same particles. The intensity of the particles in these maps gives an indication of the elemental concentration. Note that the quartz grains in Fig. 1a (some are indicated by arrows) are dark in color in Fig. 1b. Only the iron containing minerals, including basaltic glass, are lighter in color. Numerous 1–10 μm and some 60 μm quartz particles were found, with a total deposit contribution of 1–3%. It is assumed that these particles originate from Asia and were carried over more than 10 000 km by late Quaternary storms.

Patterson et al. [20] analyzed remote continental and marine aerosols, collected

with an aircraft at altitudes ranging from the air–sea exchange zone to the free troposphere. For the continental aerosols no significant differences were noted as a function of altitude. The aerosol mass with a diameter greater than 0.5 μm mainly consisted of soil dust particles, while the sulfate particles dominated the small fraction (diameter <0.5 μm). However, the composition of the samples from the Pacific showed large variations with altitude. Also noticeable was a decrease in the crustal components from the northern towards the southern Pacific.

Individual aerosol particles were sampled over the Amazon Basin and analyzed using automated EPXMA [13, 15, 21]. The results were used to estimate the chemical processes occurring in the Amazon Basin atmosphere and the aerosol and gas emission of the forest. The two main components, soil dust and biological derived material, could be linked to local sources.

Pacific aerosols were analyzed with manual and automated EPXMA. Manual analysis of central Pacific aerosols revealed that a large amount of aluminosilicates was present as an internal mixture with sea salt aerosols [22]. These mixtures are probably the result of a coagulation of aluminosilicates and sea salt particles inside the clouds. Automated analysis of more than 5000 individual particles from the eastern Pacific showed that more than 45% of all particles were 'pure' sulfur (no other detectable elements) [23]. The morphology and the diameter of these particles were used to discriminate between submicrometer particles which evaporated under the electron beam and which are most probably ammonium sulfate and less volatile particles with a mean diameter of 2 μm and more spherical in shape. A considerable portion of the particles has only characteristic calcium and phosphorus X-ray signals. Their abundance seemed to show some increase as the sampling location approached the continent, but no final conclusions could be drawn.

For several years automated EPXMA has been used to characterize individual North Sea aerosols [24]. During a first research period, 2500 particles sampled from a research vessel were sized, chemically analyzed and classified [25]. Sea salt particles dominate in air masses which originate from above the Atlantic Ocean. When the air has been above land for a considerable time, the amounts of aluminosilicates (mostly spherical fly ash particles), carbonates, gypsum, and spherical iron oxides increase. During the following 4 years more than 25 000 particles from over the North Sea and the English Channel were collected on board a research vessel [26]. The combination of single-particle analysis and hierarchical clustering revealed that the majority of the North Sea aerosols are composed of sea salt, sulfur-rich particles, silicates, and gypsum. Meteorological conditions and sampling locations influence the abundance of the particle types. A study of the difference between the samples was based on principal component analysis (PCA) of the abundances of the different particle types. A division in components was found and all could be source appointed. The marine derived aerosols were the first group, and their importance grows with increasing wind speed (sea salt aerosol formation is dominated by the breaking of waves, which is dependent on wind speed) and sampling at more marine

locations. Anthropogenic derived gypsum, which seemed to be of more importance when the air originated from above southern England, formed the second component. The samples in the third group were characterized by particles with a high silicate and sulfur content. A second PCA differentiated between the two sources of this last group, namely a mixed marine/continental origin and a truly continental one.

The aerosol composition at different heights above the North Sea was studied by the automated EPXMA of 50 000 particles which were collected at six different altitudes [27]. The continental air masses were dominated by aluminosilicates, gypsum, and iron rich particles. Due to the turbulence of the atmosphere no major difference in main particle types with altitude were noted for either western or marine air masses. Aluminosilicates, iron rich particles, sea salt enriched with lead, and organic materials were the major particle types in these air masses. Combustion or energy-generation processes seemed to be responsible for about 60% of the analyzed particles.

The modification of particle composition by air–sea interactions was examined by analyzing aerosol and rain water samples, which were collected on two research vessels located on the diameter of a 200 km circle one downwind from the other in the central part of the North Sea [28]. Automated EPXMA in combination with hierarchical clustering revealed three to eight different particle types and a principal factor analysis (PFA) appointed these groups to four major sources. As the air is crossing the North Sea the quantity of aluminosilicate particles is reduced and replaced by sea salt and sea-water crystallization products. No shrinking of the particle diameter was observed. Manual EPXMA was used to study the relation between particle composition, origin, and shape. For most particle types a shape-based characterization was possible. The collected rainwater showed serious variation in composition and diameter.

Giant aerosols have a diameter $>1\,\mu m$. The amount of these particles in the lower troposphere is small compared with the condensation-mode particles, but their contribution to atmospheric deposition is of great importance [29–31]. A slow realization of the importance of these particles, the sampling difficulties and the measurement errors have made the giant aerosol a new and still little explored research domain. The importance of these giant aerosols in the North Sea atmosphere was tested using samples which were collected on an impactor rod on top of an aircraft [32]. These samples were analyzed by a combination of EPXMA with different multivariate techniques. PFA revealed aluminosilicate dispersal, combustion processes, industrial activity, and marine origin as the major sources for these giant aerosols. Samples associated with marine and continental air masses could be differentiated by means of clustering analysis. The size distribution of the main particle types fitted a log–normal size distribution curve well, with an average diameter of about $3\,\mu m$. Only aluminosilicates had a bimodal size distribution and average size maxima at 4 and $15\,\mu m$.

Studies of volcanic ash particles using a combination of SEM and X-ray photoelectron spectroscopy (XPS) have been reported by several researchers [33–35]. The EDX spectra of these particles show major aluminum, calcium, potassium, and

silicon peaks, and small contributions for iron, manganese, and titanium. Rampino and Self [36] have proven that these volcanic aerosols are a fraction of the terrestrial particles released directly into the stratosphere. Other sources of stratospheric particles are sulfuric acid nucleation, sapphires, and meteorites [37]. The major particle types are aluminum metal particles and Al_2O_3 spheres. Sulfate particles of terrestrial origin dominate the submicrometer aerosol mass [38]. Volcanic eruptions also emit large amounts of sulfur. The thin sulfate coating on ash particles is probably the result of reaction with this sulfur in stratospheric clouds [39].

The influence of the huge oil fires in Kuwait in 1991 on the local and global environment was studied by analyzing aerosol samples collected on board an aircraft at different altitudes in the plumes [40]. Analysis of transported smoke clouds revealed more small sulfate particles, which are most probably produced by the oxidation and nucleation of sulfur oxide. In areas where rarefication of the plumes has occurred, large soot, salt, and dust particles are coated with sulfur and become active cloud condensation nuclei (CCN). The CCN initiate the formation of cloud, haze, fog, and smog droplets. In the absence of rain showers, black carbon particles, such as those formed by these fires, have a tendency to remain airborne and can undergo long-range transportation. Air trajectory studies show that there were periods when the aerosols generated by the oil fires could have reached China. The relation between the fires in Kuwait and the severe rainfall in China at the end of May and the beginning of June 1991 will have to be proven with more research.

Individual stack fly ash particles from a Hungarian power station [41] and Hungarian background aerosols [42] have both been analyzed using EPXMA. In the power-plant aerosol fraction of particles $<2\,\mu m$, two unexpected particle types were found: barium rich particles and arsenic rich gypsum particles. Anthropogenic sources also seemed to be the major origin of the $0.3–20\,\mu m$ background aerosols. Air back-trajectories confirm this statement.

The urban aerosol composition is highly variable and depends on geographical location and local activities and industries. Aerosols from the city of Khartoum, Sudan, have been analyzed using different bulk techniques and automated EPXMA for single-particle analysis [43]. The majority of these aerosols were identified as soil dust, namely aluminosilicates and smaller amounts of quartz and calcium carbonate. This enormous amount of natural airborne dust has, to our knowledge, the lowest concentration of sulfur, chlorine, zinc, bromine, and lead in urban aerosols ever reported. The daily variation in the concentration of bromine, lead, and crustal elements indicates that the production is from traffic. Aerosol particles were also collected 4 km west of Santiago, Chile, and analyzed using automated EPXMA [44]. Both the coarse and the fine fraction could be classified into eight particle types. The coarse-particle fraction was dominated by soil dust particles, while anthropogenic particles were predominant in the fine mode. Although a large amount of particles contain sulfur, they are not all of one particle type but are present in six groups. The automated EPXMA of 15 000 aerosols from Antwerp, Belgium, showed that soil dust dominates these samples [45, 46]. Sulfates, lead-rich particles, and other

anthropogenic particles are also often detected. The combustion of leaded fuel is the main source of lead containing particles. Post and Buseck [47] used SEM analysis to characterize 8000 individual urban aerosols from Phoenix, Arizona. The coarse fraction consisted of crustal material and a minor fraction of biological material, sulfur rich and lead rich particles. Some 60–80% of the fine fraction were sulfur containing particles, presumably ammonium sulfates.

Hunt et al. [48] used SEM-EDX to study lead containing particles in the floor dust of 16 houses in the London borough of Richmond, England. The identification of the different particle types was based on a classification scheme which was constructed by analyzing different types of lead source particles [49]. Lead rich particles with diameters ranging from 0 to 64 μm were identified as paint, road dust, and garden soil. Paint was also the major contributor to the lead containing dust in the size range 64–100 μm. The contribution of the major sources seemed to be independent of the age of the houses.

A combination of SEM and transmission electron microscopy (TEM) with thin-film chemical tests was used to determine the nature and morphology of methane sulfonic acid (MSA) particles, which were collected over Sakushima Island, Japan [50]. Dimethyl sulfate, which is transferred from the sea surface into the atmosphere, is oxidized to MSA and sulfur dioxide. MSA containing particles dominate in the coarse fraction of very humid air masses. These mixed particles are produced by heterogeneous nucleation reactions of gaseous MSA with sea salt and soil particles. The submicrometer fraction is dominated by sulfate particles and contains no MSA particles.

Pardess et al. [51] have recently developed a new method for determining the sulfur content in single particles by SEM. Artificially generated sulfur particles of known composition were used to calculate the calibration curves. These curves give a reasonably accurate determination of the sulfur mass in particles larger than 0.4 μm and with a minimum of 10^{-13} g of sulfur. This method can also be used to estimate the sulfur contents of heterogeneous particles.

Industrial and workplace aerosols, especially fly ash, are often analyzed using EPXMA and SEM [52–63]. The term 'fly ash' is used for a variety of particles which are emitted during combustion processes. Both the micrometer and the submicrometer size particles have silicon, aluminum, potassium, iron, titanium, manganese, and sulfur as their major elements, and minor contributions from calcium, phosphorus, sodium, chlorine, and nickel. The fly ash particles are known to have a characteristic spherical shape, although irregularly shaped particles were also observed [63]. Power plants running on oil or coal produce different fly ashes [64–67]. Oil fly ash particles contain considerably more sulfur and substantial concentrations of vanadium and nickel. Over 90% of the particles contribute to the fine fraction. The long exposure to heat and oxidants make their morphology variable from nearly spherical to lacy and spongy lumps [68]. These spongy structures easily break down into smaller aggregates. Over 90% of coal fly ash is in the coarse particle fraction [66]. They have a smooth mineral spherical morphology and contain less cenospheres.

4.2.3 Applications to Suspensions and Sediments

Sedimentation processes, interactions between the water column and the sediments, and physicochemical reactions of suspended and sedimented particles have been studied extensively using EPXMA. Initially, manual SEM-EDX was used for the chemical and morphological characterization of aquatic particles [69–72]. Bishop, Biscaye, and coworkers [73, 74] were the first to report results of automated EPXMA on aquatic particles. They analyzed individual particles from the nepheloid layer of the Atlantic Ocean and classified them according to their silicon/aluminum ratio.

EPXMA and SEM were used to identify the sources and study the lateral distribution of suspended matter collected around Sumbawa Island, Indonesia [75]. Iron rich particles, probably of volcanic origin, dominate samples collected from the northern side of the island. Suspended matter from Saleh Bay, a huge bay on the Sumbawa Island near to the Tambora volcano, is characterized by a high content of tin rich particles. Considerable amounts of tin rich particles have also been found in the Flores Sea and the Makasar Strait [76, 77]. There are possible regional sources of tin, but other sources cannot be excluded. A morphological study of aggregates showed that these exist only in the presence of biogenic material such as diatoms and coccolithophorids. The correlation between aggregates and the presence of diatoms has been reported previously [76, 78]. The flocculated particles are probably held together by a gummy organic material produced by the diatoms [75].

North Sea bulk water and corresponding surface microlayer samples have been characterized by a combination of EPXMA and laser microprobe mass spectrometry (LMMS) [79]. Due to the roughness of the sea, only small differences in particle type and abundance could be found between the collected surface microlayer and the underlying bulk samples. Organic particulate matter was more abundant in the microlayer and showed a correlation with the metal-rich fraction. Figures 2 and 3 show the secondary electron image and the X-ray spectrum of a fly-ash particle (aluminosilicate rich in iron, calcium, manganese, and sulfur) and a diatom (a silicon rich skeleton with a small aluminum fraction), both found in the surface microlayer of the North Sea. These marine samples were also compared with the corresponding aerosol samples. Since no chemical analogs for calcium rich aluminosilicates were observed above the North Sea, they must be of aquatic origin. Titanium rich particles, however, have a similar marine and atmospheric fraction and are thus deposited from the air.

Silty and clayey sediment samples from the Baltic Sea were analyzed with automated EPXMA and conventional bulk techniques [80, 81]. Automated EPXMA showed nine particle types which could be identified as different minerals. Quartz and aluminosilicates were the most abundant particle types, but they are not characteristic for the geochemistry of the Baltic sediments. Anoxic sediment environments, such as the Gotland Deep, are characterized by reduced species, like iron sulfides and rhodochrosite. Iron and manganese oxides/hydroxides indicate oxic environments such as Kattegat. In regions with periodic redox changes, sulfides and

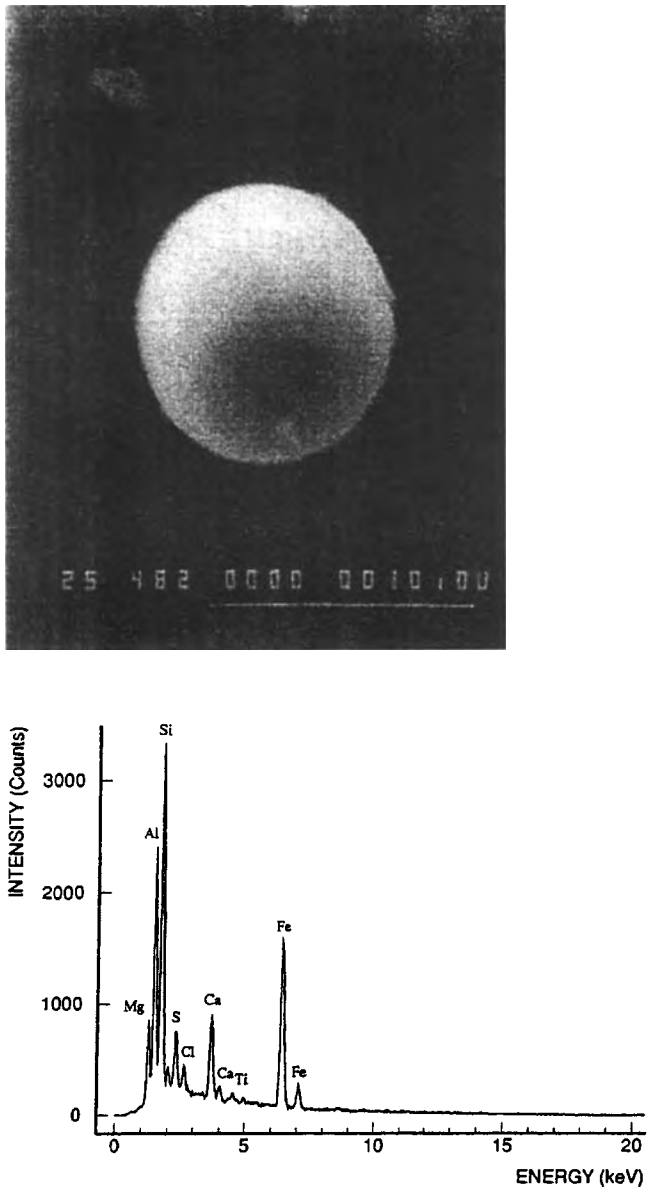


Figure 2. Secondary electron image and X-ray spectrum of a fly-ash particle collected from the surface microlayer of the North Sea [79]. Reproduced by permission of the American Chemical Society.

oxides/hydroxides are both present. Calcium rich particles are only detected in the transition zone between the Baltic Sea and the North Sea. The clear gradient from the North Sea to the Baltic Sea indicates the inflow of calcium rich particles from the North Sea [80]. The high abundance of titanium rich particles near the coast of Germany and southern Poland could be the result of erosion or of anthropogenic deposits [81].

Automated EPXMA has also been used to study deposition, remobilization, and mixing of materials from different origins in estuaries. Analysis of suspended matter from the Ems estuary (at the border of

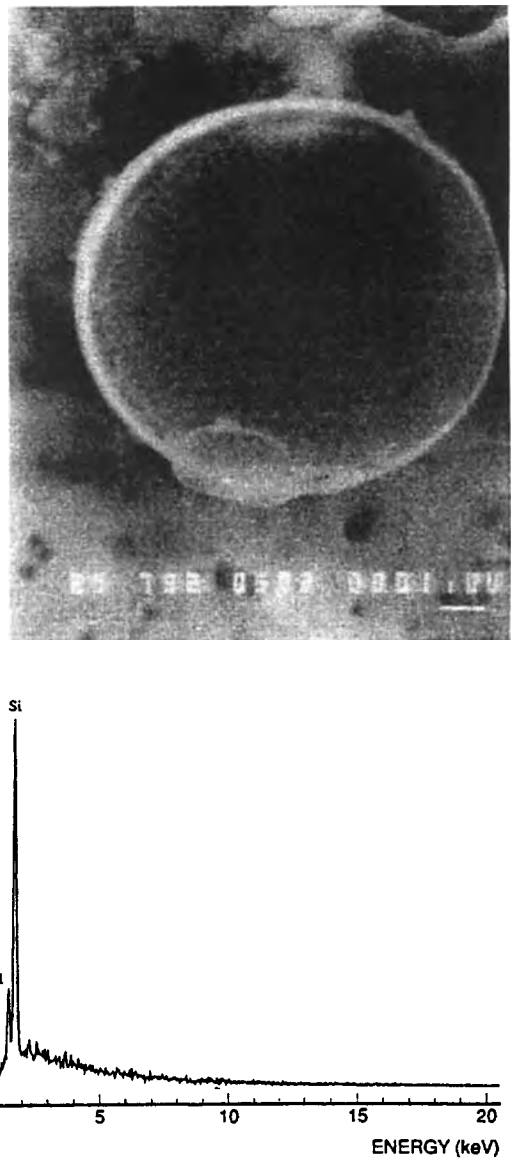


Figure 3. Secondary electron image and X-ray spectrum of a diatom skeleton present in the surface microlayer of the North Sea [79]. Reproduced by permission of the American Chemical Society.

Germany and The Netherlands) revealed mixing with marine material in the freshwater zone and the transport of this marine suspension across the salt wedge [82]. No net flux of marine suspended matter into the estuaries of the Garone or Rhône (France) was found [83, 84]. This is probably due to the different natures of these estuaries. Sediment samples from a longitunal profile and suspended matter from a maximum peak flow were collected in the Elbe river (Germany) [85]. The mixing ratio of marine and fluvial sediments could be based on the variation in abundance of calcium carbonate, and the results are comparable with those found

for the Ems river. Load transport, characteristic of discharge events, and erosion of settled mud particles are the major sources of suspended matter transported during peak events.

The nature of the inorganic colloidal and suspended matter transported by the Magela Creek, northern Australia, was studied during a small flood event [86]. Only a small difference was noticed between the inorganic composition of the coarse colloidal matter (0.1–1 μm) and the suspended particulate matter (>1 μm). Both are dominated by weathering products.

Computer-controlled SEM (CCSEM) was used to analyze suspended matter from Onondaga Lake, New York [87]. The size and composition of these particles are directly related to biological production, precipitation in the water column, and transport of suspended matter by the lake tributaries (especially after runoff events). The total amount of suspended matter and the relative abundance of the different particle types are exposed to strong temporal variations.

Organic particles can be investigated using automated SEM-EDX analysis, by staining the samples with ruthenium tetroxide. However, in EDX spectra serious interference occurs between the ruthenium L_α peak and the chlorine K_α peak. This interference is not limiting for aquatic samples, because chlorine is usually removed during their preparation. Suspended samples from the Sargasso Sea were prepared using this stain [88]. Analysis showed that over 70% of the particles were organic and the major components of the inorganic fraction were iron, calcium, sulfur, and silicon rich particles.

The correlation between the occurrence of suspended and sedimented barite and the intensity of previous biological activity [89] was confirmed by manual EPXMA of suspended matter from the Scotia Sea–Weddel Sea confluence [90]. In the surface waters, barite is mainly contained within large bioaggregates. These aggregates disperse in the subsurface zone and the barite crystals become individual particles. No association of barite with phytoplankton was found. The pelagic barite is thus most probably formed through passive precipitation in a saturated microenvironment. Manual EPXMA has also been used to distinguish the different structures of manganese [91] and the different species of pyrite [92].

The morphological characterization of environmental particles with SEM has been reported by several researchers [75, 93–95].

4.3 Scanning Transmission Electron Microscopy

Publications on the characterization of environmental particles with scanning transmission electron microscopy–energy dispersive X-ray detection (STEM-EDX) are limited. To our knowledge no articles on aerosol research have been published.

Recently, a study has been started to characterize individual submicrometer indoor aerosol particles using STEM-EDX [96]. The aim of this study is to identify those aerosol particles which are responsible for the staining of paintings in the Correr Museum, Venice. Preliminary results show the presence of silicon and sulfur rich particles, titanium rich particles, and aluminosilicates.

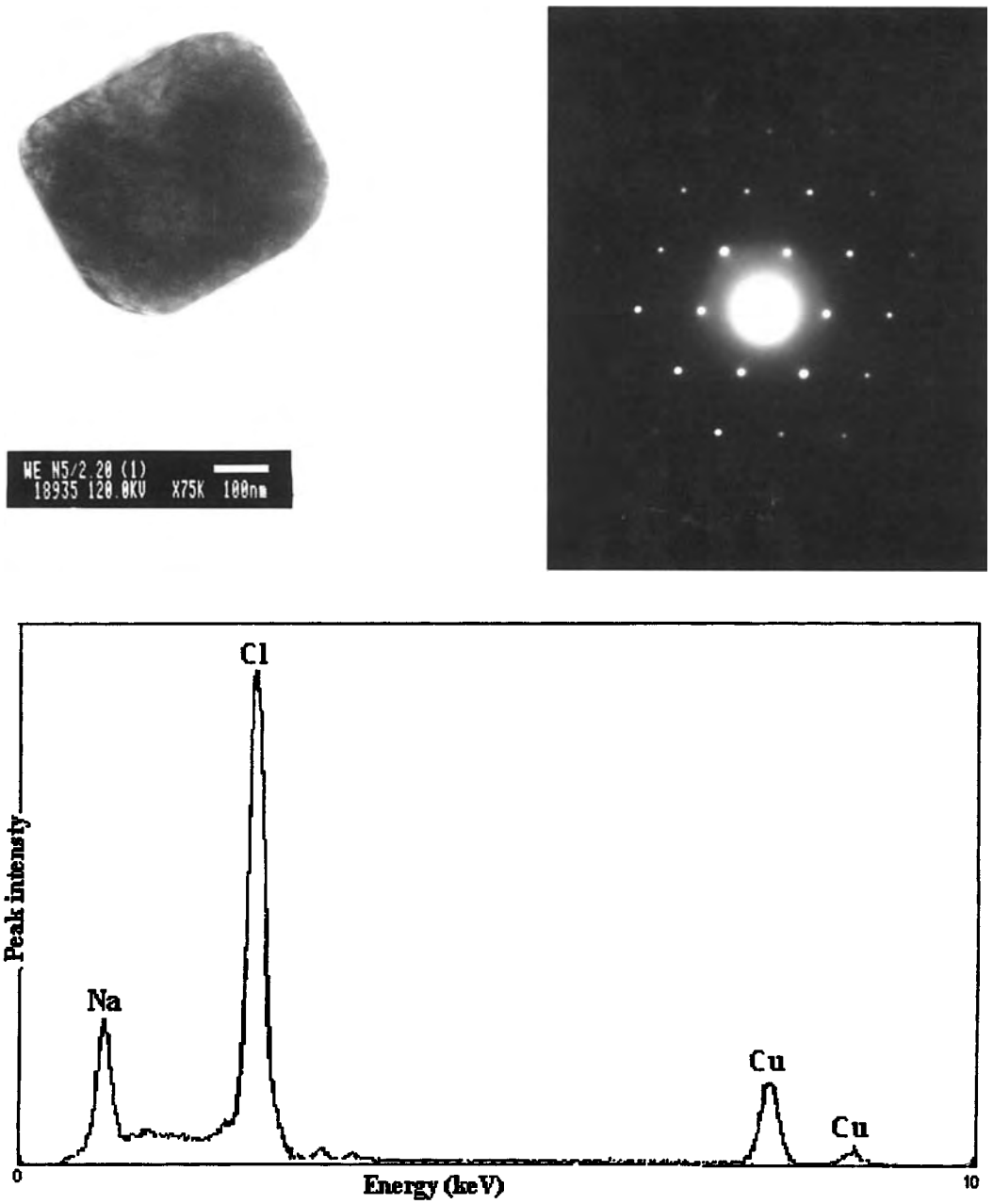


Figure 4. Transmission electron image (a), electron diffraction pattern (b), and X-ray spectrum (c) of a sodium chloride particle (halite) collected in the Southern Bight of the North Sea.

Leppard [97] has reviewed the application of electron microscopy (EM) to aquatic colloids. STEM-EDX was used for the characterization of mineral and mineral–organic colloids. The crucial point of the EM analysis of aquatic particles is sample preparation. Simple and powerful preparation techniques which reduce the amount of artifacts are described by Leppard [97] and Perret et al. [98].

Iron oxyhydroxides in the sediments of eight lakes were analyzed after preparation with Nanoplast resin [99]. A combination of STEM-EDX with electron diffraction (ED) and X-ray diffraction revealed that crystalline iron(III) oxyhydroxides were only present as poorly ordered ferrihydrite and lepidocrocite. Deposition of these iron oxyhydroxides in the lake sediments also seemed to occur on bacterial cells and their exopolymers. This explains the significant amount of organic carbon found in the diagenetic iron(III) oxyhydroxides.

Identification of the mineral structure of crystalline particles is possible by combining STEM-EDX with selected-area ED. This combination has been tested on Nanoplast-embedded micrometer-sized suspended particles collected in the Southern Bight of the North Sea, and the results are promising [100]. As a preliminary result of this research the transmission electron image, electron diffraction pattern, and X-ray spectrum of a sodium chloride (halite) particle are shown in Fig. 4.

4.4 Electron Energy-Loss Spectrometry

Electron energy-loss spectrometry (EELS) has only been recently commercialized and for environmental applications it is still in an experimental stage. The few applications published on this subject are briefly discussed.

Standard test aerosols of inorganic salts were analyzed by serial and parallel EELS (SEELS and PEELS) [101]. Even at cryogenic temperatures there were distinct losses and structure reorganizations of the inorganic salts. PEELS can be used to analyze these beam-sensitive samples, but only for qualitative purposes.

PEELS analysis of atmospheric aerosols indicated that the majority of the ultrafine particles consist of carbon [102]. The carbonaceous aerosols collected in Phoenix, Arizona, are smaller than 2 μm and contain 10–100 aggregated spheres [103, 104]. Information on the electronic state of the carbon edge showed the presence of a mixture of graphitic and amorphous carbon within and among the individual aggregate particles. The graphite is probably part of the primary soot, while the amorphous carbon is the result of hydrocarbon condensation during aggregation. During the summer, some particles are coated with sulfates and nitrates. Visual evidence of these sulfur coatings was provided by element-specific imaging (ESI) [104]. Since coated and uncoated aggregates have the same structural variations, it is presumed that they are from the same source.

The combination of EELS with ESI is very useful in studying the inhomogeneous composition of nanometer-range, surface-modified asbestos fibers [105, 106]. Carbon maps or organosilane coated chrysolite fibers showed that the coating was not distributed homogeneously over the fibers. Some fibers even seem to be unaffected by the treatment. The surface modification,

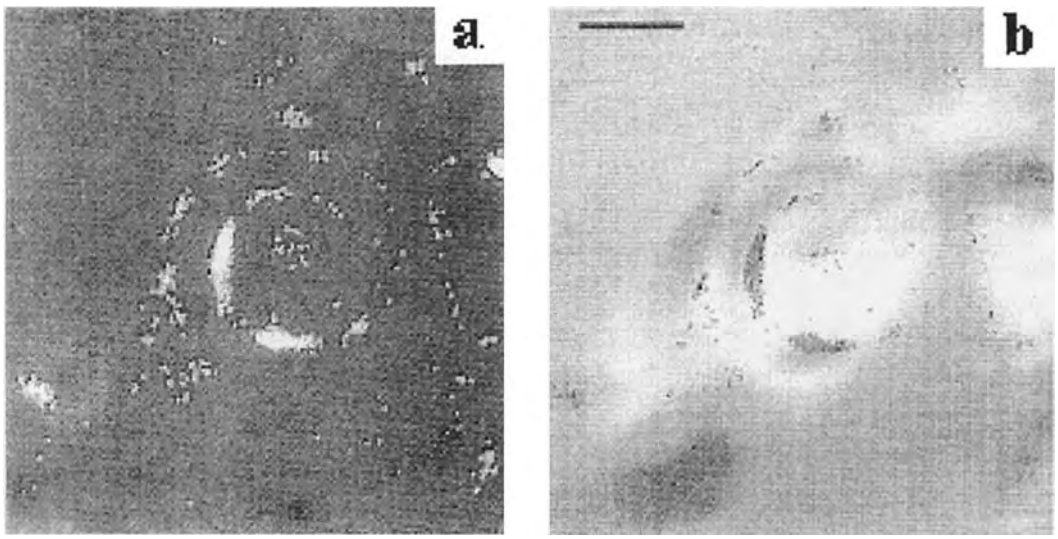


Figure 5. Element-specific imaging of titanium in a cross-section of a $TiCl_3$ treated asbestos fiber. (a) Net titanium image obtained by subtracting an extrapolated image at 460 eV (calculated from the 435 and 445 eV images) from the 460 eV image. (b) Combination of the binary titanium distribution image (black) and the grey reference electron specific image at 435 eV (scale bar = 25 nm) [106]. Reproduced by permission of Elsevier Science Publishers B.V.

necessary for the reduction of the toxicity of the fibers, is thus only partially efficient. The oxygen peaks from $TiCl_3$ modified fiber spectra have a pre-edge which indicates bonding between oxygen and titanium. The titanium chloride is thus chemically bound to the fiber. ESI maps of titanium reveal the encapsulation of the material and thus full coverage of the fibers. Elemental mappings of titanium in a cross-section of a $TiCl_3$ treated fiber embedded in Spurr resin are shown in Fig. 5.

4.5 Micro-Raman Spectrometry

Although micro-Raman spectrometry is a very slow analytical method, it offers great potential for environmental particle analysis. Environmental researchers are fairly unfamiliar with this technique and its possibilities, which has resulted in a very small amount of publications on environmental applications.

Raman microprobe measurements on Antarctic aerosols showed that the sulfur rich fine fraction is dominated by H_2SO_4, $(NH_4)HSO_4$, $(NH_4)_2SO_4$, or a mixture of more complex species [107]. The particulate aerosol fraction from oil and coal power plants has also been examined [108, 109]. Vanadium oxide was found to be a major component of the oil-derived particles, and only a minor one in the coal ashes.

A combination of Raman and infrared spectrometry was used to characterize large individual dust particles from an office–laboratory area [110]. Not surprisingly, many of the analyzed dust particles originated from paper products.

4.6 Fourier Transform Infrared Spectrometry

FT-IR spectroscopy and FT-IR microspectroscopy have rarely been used for environmental applications. Long-path FT-IR spectroscopy has been reported for the measurement of atmospheric trace gases [111], the monitoring of airborne gases and vapours [112–114], the study of atmospheric pollutants [115], and the characterization of organic components in aerosols [116].

FT-IR microspectroscopy has been used as a direct method to analyze the organic contents of size-fractionated atmospheric aerosols [117] and size-segmented aerosol particles and clusters [118]. In these publications it is stated that FT-IR microspectroscopy has three major advantages for the analysis of single aerosol particles: only small sample masses are required, sample dilution is not necessary, and direct analysis of the impactor stages after aerosol collection is possible. These statements were tested by the analysis of giant North Sea aerosols with FT-IR microspectroscopy and EPXMA [96]. The first results were not promising. The analyzed particles were dominated by inorganic salts which only absorb partially in the middle IR region. Most marine giant aerosols contain a mixture of components which leads to peak overlap in the spectra and to interpretation difficulties.

Acknowledgements

This work was partially prepared in the framework of the Impulse Programme in Marine Sciences, supported by the Belgian State–Prime Minister's Service–Services for Scientific, Technical and Cultural Affairs (contract MS/06/050).

4.7 References

[1] M. Grasserbauer, *Mikrochim. Acta* **1983,** *III*, 415.
[2] R. Van Grieken, P. Artaxo, P. Bernard, L. Leysen, P. Otten, H. Storms, A. Van Put, L. Wouters, C. Xhoffer, *Chemica Analitycza* **1990,** *35*, 75.
[3] R. Van Grieken, C. Xhoffer, *J. Anal. Atom. Spectrom.* **1992,** *7*, 81.
[4] C. Xhoffer, L. Wouters, P. Artaxo, A. Van Put, R. Van Grieken in *Environmental Particles, Vol. I* (Eds.: J. Buffle, H. P. Van Leeuwen), Lewis, Chelsea MI **1992,** Chap. 3.
[5] W. Jambers, L. De Bock, R. Van Grieken, *Analyst* **1995,** *120*, 681.
[6] R. Cadle, W. Fisher, E. Frank, J. Lodge, *J. Atmos. Sci.* **1967,** *25*, 100.
[7] F. Parungo, B. Bodhaine, J. Bortnak, *J. Aerosol Sci.* **1981,** *12*, 367.
[8] G. E. Shaw, *Atmos. Environ.* **1983,** *17*, 329.
[9] T. Ito, *J. Atmos. Chem.* **1985,** *3*, 69.
[10] B. A. Bodhaine, M. E. Murphy, *J. Aerosol Sci.* **1980,** *11*, 305.
[11] B. A. Bodhaine, J. J. Deluisi, J. M. Harris, *Tellus* **1986,** *38B*, 223.
[12] P. Artaxo, M. L. C. Rabello, W. Maenhaut, R. Van Grieken, *Tellus* **1992,** *44B*, 318.
[13] P. Artaxo, R. Van Grieken, F. Watt, M. Jakšić in *Proc. Second World Congress on Particle Technology*, Society of Powder Technology, Kyoto **1990,** pp. 421–426.
[14] R. E. Van Grieken, P. Artaxo, C. Xhoffer in *Proc. Fiftieth Annual Meeting by the Electron Microscopy Society of America* (Eds.: G. W. Bailey, J. Bentley, J. A. Small), San Francisco Press, San Francisco, CA **1992,** pp. 1482–1483.
[15] H. Storms, *Ph.D. Thesis,* University of Antwerp, Antwerp **1988.**
[16] E. Bigg, *J. Appl. Meteorol.* **1980,** *19*, 521.
[17] W. Cunningham, W. Zoller, *J. Aerosol Sci.* **1981,** *12*, 367.
[18] H. Van Malderen, R. Van Grieken, T. V. Khodzher, N. S. Bufetov, K. P. Koutsenogii, *Atmos. Oceanic Opt.* **1994,** *7*, 622.
[19] E. J. Begét, M. Keskinen, K. Severin, *Sediment. Geol.* **1993,** *84*, 189.
[20] E. M. Patterson, C. S. Kiang, A. C. Delany, A. F. Artburg, A. C. D. Leslie, B. J. Huebert, *J. Geophys. Res.* **1980,** *8*, 7361.

[21] P. Artaxo, H. Storms, F. Bruynseels, R. Van Grieken, W. Maenhaut, *J. Geophys. Res.* **1988,** *93*, 1605.
[22] M. O. Andreae, R. J. Charlson, F. Bruynseels, H. Storms, R. Van Grieken, W. Maenhaut, *Science* **1986,** *232*, 1620.
[23] C. Xhoffer, *M.Sc. Thesis*, University of Antwerp, Antwerp **1987.**
[24] H. Van Malderen, L. De Bock, J. Injuk, C. Xhoffer, R. Van Grieken in *Progress in Belgian Oceanographic Research*, Royal Academy of Belgium, Brussels **1993,** pp. 119–135.
[25] F. Bruynseels, H. Storms, R. Van Grieken, L. Van Der Auwera, *Atmos. Environ.* **1988,** *22*, 2593.
[26] C. Xhoffer, P. Bernard, R. Van Grieken, L. Van Der Auwera, *Environ. Sci. Technol.* **1991,** *25*, 1470.
[27] C. M. Rojas, R. Van Grieken, *Atmos. Environ.* **1992,** *26A*, 1231.
[28] L. A. De Bock, H. Van Malderen, R. Van Grieken, *Environ. Sci. Technol.* **1994,** *28*, 1513.
[29] R. Jaenicke, *Ann. NY Acad. Sci.* **1980,** *338*, 317.
[30] R. Jaenicke in *Chemistry of the Unpolluted and Polluted Troposphere*, 1st edn. (Eds.: W. Georgii, W. Jaeschke), D. Reidel, Dordrecht **1982,** pp. 341–374.
[31] H. L. Dedeurwaerder, *Ph.D. Thesis*, Free University of Brussels, Brussels **1988.**
[32] H. Van Malderen, C. Rojas, R. Van Grieken, *Environ. Sci. Technol.* **1992,** *26*, 750.
[33] J. Fruchter, *Science* **1980,** *209*, 1116.
[34] J. P. Wightman, *Colloid. Surf.* **1982,** *4*, 401.
[35] J. L. Gooding, U. S. Clanton, E. M. Gabel, J. L. Warren, *Geophys. Res. Lett.* **1983,** *10*, 1033.
[36] M. R. Rampino, S. Self, *Sci. Am.* **1984,** *250*, 48.
[37] D. E. Brownlee, *Proc. Annual Conference on Microbeam Analysis Society* **1990,** *15*, 199.
[38] R. D. Cadle, G. W. Grams, *Rev. Geophys. Space Phys.* **1975,** *13*, 475.
[39] I. D. R. Mackinnon, J. L. Gooding, D. S. McKay, U. S. Clanton, *J. Volcanol. Geotherm. Res.* **1984,** *23*, 125.
[40] F. Parungo, B. Kopcewicz, C. Nagamoto, R. Schnell, P. Sheridan, C. Zhu, J. Harris, *J. Geophys. Res.* **1992,** *97*, 15867.
[41] S. Sándor, S. Török, C. Xhoffer, R. Van Grieken in *Proc. Twelfth International Congress for Electron Microscopy* (Eds.: L. D. Peachy, P. B. Williams), San Francisco Press, San Francisco, CA **1990,** pp. 245–255.
[42] S. Török, S. Sándor, C. Xhoffer, R. Van Grieken, E. Mészáros, A. Molnar, *Idojaras* **1993,** *96*, 223.
[43] M. A. H. Eltayeb, C. F. Xhoffer, P. J. Van Espen, R. E. Van Grieken, *Atmos. Environ.* **1993,** *27B*, 67.
[44] C. M. Rojas, P. Artaxo, R. Van Grieken, *Atmos. Environ.* **1990,** *24B*, 227.
[45] W. Van Borm, F. Adams, W. Maenhaut, *Atmos. Environ.* **1989,** *23*, 1139.
[46] W. Van Borm, L. Wouters, R. Van Grieken, F. Adams, *Sci. Total Environ.* **1990,** *90*, 55.
[47] J. E. Post, P. R. Buseck, *Environ. Sci. Technol.* **1985,** *18*, 35.
[48] A. Hunt, D. L. Johnson, I. Thornton, J. M. Watt, *Sci. Total Environ.* **1993,** *138*, 183.
[49] A. Hunt, D. L. Johnson, J. M. Watt, I. Thornton, *Environ. Sci. Technol.* **1992,** *26*, 1513.
[50] G.-W. Qian, Y. Ishizaka, *J. Geophys. Res. Oceans* **1993,** *98*, 8459.
[51] D. Pardess, Z. Levin, E. Ganor, *Atmos. Environ.* **1992,** *26A*, 675.
[52] G. Bonafede, L. T. Kiss, *Am. Soc. Mech. Eng.* **1973,** 73-WA/CD-7.
[53] J. E. Baker, C. A. Evans, A. Loh, D. F. S. Natush, *Proc. Annual Conference on Microbeam Analysis Society* **1975,** *10*, 33A.
[54] D. L. Gibbon, *Scanning Electron Microsc.* **1979,** *I*, 501.
[55] R. L. Carpenter, R. D. Clark, Y.-F. Su, *J. Air Pollut. Control Assoc.* **1980,** *30*, 679.
[56] D. Lichtman, S. Mroczkowski, *Environ. Sci. Technol.* **1985,** *19*, 274.
[57] L. M. Middleman, J. D. Geller, *Proc. Annual Scanning Electron Microscopy Symposium* **1976,** 171.
[58] J. A. Small, W. H. Zoller, *NBS Monogr. (US)* **1977,** *464*, 651.
[59] G. L. Fisher, B. A. Prentice, D. Silberman, J. M. Ondov, *Environ. Sci. Technol.* **1978,** *12*, 447.
[60] F. Parungo, E. Ackerman, H. Proulx, R. Pueschel, *Atmos. Environ.* **1978,** *12*, 929.
[61] T. L. Hayes, J. B. Pawley, G. L. Fisher, *Scanning Electron Microsc.* **1978,** *I*, 239.
[62] R. Capron, P. Haymann, F. Pellerin, *C.R. Hebd. Seances Acad. Sci.* **1979,** *C289*, 313.
[63] N. Kaufherr, D. Lichtman, *Environ. Sci. Technol.* **1984,** *18*, 544.
[64] J. Wagman in *Colloid and Interface Science, Vol. II: Aerosols, Emulsions and Surfactants*, Academic Press, New York **1976.**
[65] E. Denoyer, T. Mauney, D. F. S. Natusch, F. Adams, *Proc. Annual Conference on Microbeam Analysis Society* **1982,** *17*, 191.
[66] Y. Mamane, J. L. Miller, T. G. Dzubay, *Atmos. Environ.* **1986,** *20*, 2125.
[67] B. Raeymaekers, *Ph.D. Thesis*, University of Antwerp, Antwerp **1987.**
[68] W. C. McCrone, J. G. Delly, *The Particle Atlas, Vol. II*, 2nd edn, Ann Arbor Science, Ann Arbor, MI **1973, p.** 543.
[69] F. Dehairs, R. Chesselet, J. Jedwab, *Earth Planet. Sci. Lett.* **1980,** *49*, 528.

[70] J. Jedwab, *Earth Planet. Sci. Lett.* **1980,** *49*, 551.
[71] J. M. Skei, S. Melson, *Estuarine, Coast. Shelf Sci.* **1982,** *14*, 61.
[72] B. N. Sundby, N. Silverberg, R. Chesselet, *Geochim. Cosmochim. Acta* **1984,** *45*, 293.
[73] J. K. B. Bishop, P. E. Biscaye, *Earth Planet. Sci. Lett.* **1982,** *58*, 265.
[74] C. E. Lambert, J. K. B. Bishop, P. E. Biscaye, R. Chesselet, *Earth Planet. Sci. Lett.* **1984,** *70*, 237.
[75] D. Eisma, A. Van Put, R. Van Grieken, *Mitt. Geol.-Paläont. Inst. Univ. Hamberg* **1992,** *70*, 137.
[76] D. Eisma, J. Kalf, M. Karmini, W. G. Mook, A. Van Put, P. Bernard, R. Van Grieken, *Neth. J. Sea Res.* **1989,** *24*, 383.
[77] A. Van Put, D. Eisma, R. Van Grieken, unpublished results.
[78] D. Eisma in *Facets of Modern Biogeochemistry* (Eds.: V. Ittekkot, S. Kempe, W. Michaelis, A. Spitzy), Springer Verlag, Heidelberg **1990,** pp. 127–146.
[79] C. Xhoffer, L. Wouters, R. Van Grieken, *Environ. Sci. Technol.* **1992,** *26*, 2151.
[80] P. C. Bernard, R. E. Van Grieken, L. Brügmann, *Mar. Chem.* **1989,** *26*, 155.
[81] F. Belmans, R. Van Grieken, L. Brügmann, *Mar. Chem.* **1993,** *42*, 223.
[82] P. Bernard, R. Van Grieken, D. Eisma, B. Hudec, *Environ. Sci. Technol.* **1986,** *20*, 467.
[83] D. Eisma, P. Bernard, J. Boon, R. Van Grieken, J. Kalf, W. Mook, *Mitt. Geol.-Paläont. Inst. Univ. Hamburg* **1985,** *58*, 397.
[84] D. Eisma, P. Bernard, G. Cadee, V. Ittekot, J. Kalf, R. Laane, J. Martin, W. Mook, A. Van Put, T. Schumacher, *Neth. J. Sea Res.* **1991,** *28*, 193.
[85] A. Van Put, R. Van Grieken, R.-D. Wilken, B. Hudec, *Wat. Res.* **1994,** *28*, 643.
[86] B. T. Hart, G. B. Douglas, R. Beckett, A. Van Put, R. E. Van Grieken, *Hydrol. Proc.* **1993,** *7*, 105.
[87] D. L. Johnson, J. Jiao, S. G. DosSantos, S. W. Effler, *Environ. Sci. Technol.* **1991,** *25*, 736.
[88] D. M. Lavoie, *Deep-Sea Res.* **1992,** *39*, 1655.
[89] J. Bishop, *Nature (London)* **1988,** *332*, 341.
[90] N. Stroobants, F. Dehairs, L. Goeyens, N. Vanderheijden, R. Van Grieken, *Mar. Chem.* **1991,** *35*, 411.
[91] J. Middelburg, G. De Large, H. Van der Sloot, P. Van Emburg, S. Sophiah, *Mar. Chem.* **1989,** *23*, 353.
[92] G. Luther, A. Meyerson, J. Krajewski, R. Heres, *J. Sediment. Petrol.* **1980,** *50*, 1117.
[93] A. Tucan, S. Pamukcu in *Proc. Mediterranean Conference on Environmental Geotechnology* (Eds.: A. B. Usmen, Y. B. Acer), Balkema, Rotterdam **1992,** pp. 109–121.
[94] E. Tipping, C. Woof, K. Clarke, *Hydrol. Proc.* **1993,** *7*, 263.
[95] J. M. Oades, A. G. Waters, *Aust. J. Soil Res.* **1991,** *29*, 815.
[96] L. A. De Bock, R. E. Van Grieken, unpublished work.
[97] G. G. Leppard in *Environmental Particles, Vol. I* (Eds.: J. Buffle, H. P. Van Leeuwen), Lewis, Chelsea, MI **1992,** Chap. 6.
[98] D. Perret, G. G. Leppard, M. Müller, N. Belzile, R. De Vitre, J. Buffle, *Water Res.* **1991,** *25*, 1333.
[99] D. Fortin, G. G. Leppard, A. Tessier, *Geochim. Cosmochim. Acta* **1993,** *57*, 4391.
[100] W. Jambers in *Progress in Belgian Oceanographic Research*, Royal Academy of Belgium, Brussels **1996,** pp. 115–117.
[101] C. Xhoffer, *Ph.D. Thesis*, University of Antwerp, Antwerp **1993,** pp. 189–204.
[102] A. D. Maynard, L. M. Brown, *J. Aerosol Sci.* **1992,** *23*, S433.
[103] C. Xhoffer, *Ph.D. Thesis*, University of Antwerp, Antwerp **1993,** pp. 205–233.
[104] K. A. Katrinak, P. Rez, P. R. Buseck, *Environ. Sci. Technol.* **1992,** *26*, 1967.
[105] C. Xhoffer, P. Berghmans, I. Muir, W. Jacob, R. Van Grieken, F. Adams, *J. Microsc.* **1991,** *162*, 179.
[106] P. Berghmans, J. Injuk, R. Van Grieken, F. Adams, *Anal. Chim. Acta* **1994,** *297*, 27.
[107] E. Etz, G. Rosasco in *Proc. 5th International Conference on Raman Spectroscopy*, Hans Ferdinand Schulz Verlag, Freiburg **1976,** pp. 776–777.
[108] E. Etz, G. Rosasco, J. Blaha in *Environmental Pollutants: Detection and Measurements* (Eds.: T. Toribara, J. Coleman, B. Dahneke, I. Feldman), Plenum Press, New York **1978,** p. 413.
[109] E. Etz, G. Rosasco, K. Heinrich, *EPA Report, EPA-600/2-78-193* **1978.**
[110] P. Lang, J. Katon, A. Bonanno, *Appl. Spectrosc.* **1988,** *42*, 313.
[111] R. A. Nyquist, M. A. Leugers, M. L. McKelvy, R. R. Papenfuss, C. L. Putzig, L. Yurga, *Anal. Chem.* **1990,** *62*, 223R.
[112] J. Gosz, C. Dahm, P. Risser, *Ecology* **1988,** *69*, 1326.
[113] S. Levine, Y. Li-Shi, C. Strang, X. Hong-Kui, *Appl. Ind. Hyg.* **1989,** *4*, 180.
[114] Y. Li-Shi, S. Levine, *Anal. Chem.* **1989,** *61*, 677.
[115] G. Small, R. Kroutil, J. Ditillo, W. Loerop, *Anal. Chem.* **1988,** *60*, 264.
[116] R. Gordon, N. Trivedi, B. Singh, *Environ. Sci. Technol.* **1988,** *22*, 672.
[117] M. Dargler, S. Burke, S. Hering, D. Allen, *Atmos. Environ.* **1987,** *21*, 1001.
[118] R. Kellner, H. Malissa, *Aerosol Sci. Technol.* **1989,** *10*, 397.

Handbook of Microscopy
S. Amelinckx, D. van Dyck,J. van Landuyt , G. van Tendeloo

5 Quantitative Hyleography: The Determination of Quantitative Data from Micrographs

5.1 Hyleography, the Imaging of Materials

Traditionally, metallurgists have called the techniques for investigating the structure of metals *metallography*. In the 21st century materials scientists will wish and need to relate the microstructure of a wide range of solid materials to their properties. The materials of interest—many of which are covered in other articles in this *Handbook*—include metals and alloys, plastics, rubbers, ceramics, semiconductors, wood, concrete, bone and biological tissue, and a new general term is needed to describe the study of their microstructures. I have proposed [1] that we should adopt a term derived from the Greek 'hyle' meaning 'the stuff of which things are made'. Daniel [2] suggested 'hyleology' for materials science, from which I have derived, totally from the Greek, 'hyleography' and hence 'hyleographer'.

The range of techniques used for hyleography is wide. Imaging techniques range from light microscopy (LM) (Chap. 1, Vol. 1 of this Handbook) and scanning electron microscopy (SEM) to transmission electron microscopy (TEM) (Chap. 4.1, Vol. 1 of this Handbook), various ion microscopies and many scanned probe methods (SPM) (Chaps. 4.2 and 7, Vol. 2 of this Handbook). Other spectroscopic techniques permit elemental or molecular analysis. Such techniques generally rely on the use of a primary 'probing' beam (e.g., light, electrons, or X-rays) and work by detecting a secondary effect (emission of light, electrons, X-rays, etc.) as a function of position, time, or a number of other variables. Since there are at least ten probes and secondary effects in use already it is not unreasonable to estimate that more than 1000 different physical techniques for the examination of solids are potentially available, although not all of these possible combinations have yet been exploited. In this contribution we cover the major steps which must be taken to extract meaningful quantitative data from a range of hyleographic techniques.

5.2 The Need for Quantification

Microstructural parameters that are widely used in assessing materials include: the size, shape, and distribution of grains;

preferred orientation; the volume fraction and uniformity of inclusions; the presence, absence, amount, and spacing of phases and components; and local and overall composition. All of these parameters are important because of their effect on engineering properties, but, additionally, quantification of atomic-scale images and compositional data are often of equal importance. Similar considerations affect the interpretation of images and spectra in the field of bioscience (and, particularly, biomedical science).

In any attempt to be genuinely quantitative, two of the most important considerations must be the repeatability and reproducibility of the measurements and their traceability to primary standards. The measurements must be accurately repeatable by any competent microscopist and must not depend on the specific skills of a single individual, or there should be some quantitative information available on the uncertainties associated with the use of different operators. This is likely to become increasingly important as microstructural and ultrastructural data are used in legal and patent proceedings. For the same reasons dating, timing and archiving of all microstructural images and spectral data will become widespread, and are already being practised in some laboratories.

5.3 Instrument Calibration

5.3.1 Calibration of Microscopes

The parameter most obviously needing calibration is the magnification of the microscope used to study the specimen although this is not always straightforward. It is quite easy to image a ruler which has been calibrated against a primary or secondary standard; however, this is rarely useful beyond a magnification of about 100×. The UK National Physical Laboratory (NPL) has prepared a linewidth standard with details in the size range 0.5 to 10 μm which provides a magnification standard for higher (optical microscope) magnifications. The standard has to be calibrated using a traveling microscope fitted with a laser interferometer, and is thus traceable back to the standard meter. It is necessary, however, to calibrate each lens combination which might be used, and to check two orthogonal directions. Further calibration of the imaging chain must then be undertaken if an image analyzer is to be used. Variations with supply voltage and temperature should be negligible with modern digital imaging systems, but will not be for older analogue systems. The essential principle is that each final image magnification must be traceable to the primary standard meter.

The same considerations apply to all types of microscope. For the scanning electron microscope (SEM), quite a range of calibration standards are now available. Several of these, such as metal meshes and populations of polystyrene spheres, are not conveniently traceable to the primary standard. However, traceable standards are becoming widely available. For example, electron beam lithographically ruled silicon wafers and scribed metal standards are now available from SIRA in the UK and NIST in the USA (via commercial retailers) with certification guaranteeing 1% accuracy. NPL is working on a

0.2 μm linewidth standard suitable for the SEM. Two ASTM standards also outline SEM calibration procedures.

In the case of the TEM, magnification calibration is not so simple. The magnification, even after calibration, can vary by as much as 1.5% as soon as instrumental conditions are altered, because of hysteresis in the magnetic lenses [3]. Standard meshes are available for low magnifications but most of them are not traceable to a primary standard. At very high magnifications, structure or lattice imaging of crystals with well-known structures by high resolution electron microscopy (HREM) (see Chap. 4, Vol. 1 of this Handbook) provides an internal standard in which one can be reasonably confident. However, it is quite difficult to calibrate intermediate magnifications (e.g., between 100 000× and 5000×) reliably.

If a stage micrometer is to be used, as may be the case for fully automated analyses by light microscopy, SEM or EPMA, then it too should be traceably calibrated in both x and y directions. The NIST/NPL calibration standards mentioned above are suitable for this purpose.

The use of an image analyzer (see Chap. 8, Vol. 2 of this Handbook) brings the potential for more serious calibration problems. In addition to calibration of the magnification it is also necessary to consider the possibility of image distortion, and the effectiveness of the edge detection algorithms. There is a serious problem in relying on calibration via a standard reference graticule. The setting of thresholds on binary test objects (with sharp boundaries between black and white regions) is not necessarily the same as that appropriate to interfaces between different shades of grey, which commonly occur in real specimens. There is no substitute for good specimen preparation (see Chap. 2, Sec. 3 of this Volume) which provides clear, sharp contrast between features which are to be distinguished, together with frequent blind cross-checking between operators using the same specimens.

5.3.2 Calibration of Spectrometers

There are three key types of calibration which are needed (in addition to any microscope calibration) before a spectroscopic instrument can be used for quantitative hyleography. These are: (i) an energy calibration (e.g., to permit element identification in EDX or EELS); (ii) a count rate or detection efficiency calibration (e.g., to permit comparison of count rates from unknown and standard or from one element and another); and (iii) calibration of any necessary correction procedures (e.g., ZAF corrections in EPMA). Each of these involves considerable skill and is covered elsewhere in this *Handbook*. At this point it is sufficient to comment that calibration (iii) is very much more difficult to standardize than (i) or (ii) for almost all hyleographic techniques.

5.4 Sample Type

One of the most important assumptions of hyleographic analysis is almost always implicit and is rarely properly considered, namely, that the sample is assumed to be 'typical'. Sometimes this is assumed because the sample is selected from a specific region (e.g., from a particular

location in an engineering component) but often there is little consideration given to this important first stage of sampling. For example, it is rarely revealed in published work whether a metallic sample came from a particular place in an ingot, or whether a semiconductor sample was from the center or edge of the wafer. These considerations assume a much greater importance in quantitative studies. To illustrate their potential significance imagine that you need to study a hard-boiled egg by taking 1 mm^3 samples. The proper procedure would need to be specified in advance and might involve cutting multiple 1 mm sections (perhaps 40 of them) and then selecting every fifth section. Each of the eight selected sections would then be cut into 1 mm lengths (about 25 from each) and every fifth length chosen. Each of the 40 chosen lengths would be diced into 25 1 mm cubes and each fifth cube would be examined. This would give 200 cubes, every one of which should be studied. The total amount of material examined would be less than 1% of the whole, but it would have been selected in an unbiased manner and the original location of every sample would be known. You would then have samples containing yolk and white in meaningful numbers. This type of deliberate, unbiased sampling is very rarely applied in the materials sciences, at any scale of microstructural examination, despite the obvious problems associated with biased or inadequate sample selection. This criticism is particularly significant for TEM studies, in which the region of the sample which is imaged is likely to be not 1% of the whole but less than 10^{-9}%!

Despite the above discussion, let us assume that appropriate volumes of properly selected, perfect, artefact-free specimen can be produced, viewed at the appropriate magnification in a perfect microscope and analyzed using a perfect image analyzer. 'Perfect' in these contexts implies 'providing the required (defined) accuracy'.

The important considerations for quantification are:

- The statistical significance of the results, as a result of the (explicit or implicit) sampling procedure.
- The validity (and necessity) of extending two-dimensional measurements to three-dimensional parameters.

5.4.1 Sampling

Above we considered the macroscopic unbiased selection of samples. After satisfying ourselves at this level, we must next consider the proper sampling of the (presumably finely-divided) microstructure [4]. If measurements are to be made then there must be at least two phases of interest present. If there is a large volume fraction of the phase to be measured, or it is easily possible to obtain a large number of uniform features within a field of view, then the sampling problem reduces to a straightforward reliance on the scientist spotting any unexpected nonuniformity. It is merely necessary to record data from regions which contain a statistically significant number of features. Problems are more serious if it is necessary to make measurements in conditions where there is a low volume fraction and/or an inhomogeneous distribution of the measured features.

Chone [5] has considered this problem in the context of inclusion content assessment in continuously cast steels. His

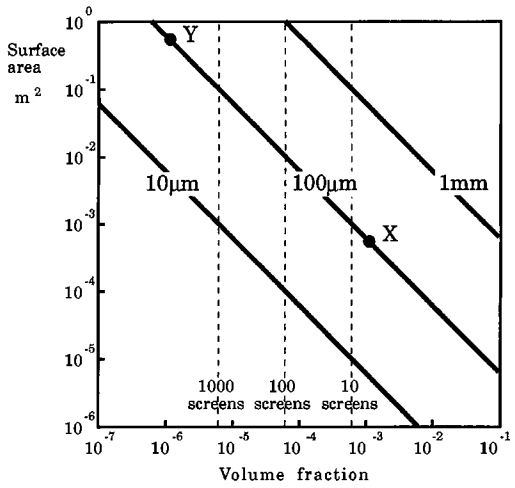


Figure 1. Chart showing the area of surface which it is necessary to examine in order to measure given features to 20% precision.

analysis was based on the assumption that the volume fraction of inclusions could be safely extrapolated from the area fraction found in a section as if they formed a dilute concentration of spherical inclusions. This may not be entirely satisfactory, but the general conclusions of his analysis are worthy of further study. Chone produced charts showing the surface area that should be examined in order to make a measurement of inclusion volume fraction with a precision of ±20%. An example is shown in Fig. 1. The three bold diagonal lines show the surface area to be examined for populations of particles of three different sizes. In calculating these lines all the particles were assumed to be the same size, which is clearly unrealistic. The general conclusions, however, are clear. For instance, to measure a volume fraction of 10^{-3} (0.1%) of 100 μm particles it is necessary to analyze 10^{-3} m^2 of specimen (the point marked X on the Figure). By contrast, to measure a volume fraction of only 10^{-6} (0.0001%), such as might easily be required in a clean sheet steel product, requires the assessment of almost 1 m^2 of specimen (point Y on the Figure).

A more easily interpreted parameter which describes the effort involved in making these measurements is the number of fields of view (or screens) which need to be analyzed in each case. If we assume that each field is examined at the appropriate magnification (that is the minimum magnification which will bring the features of interest to measurable size) then we can plot vertical lines which show the number of screens needed. Three of these are shown dotted on Fig. 1. We can now see that the measurement at point X needs fewer than 10 screens, while that at point Y needs several thousand. The conclusion is clear: the measurement of small volume fractions requires many screens and therefore takes a long time. It is almost certainly not feasible to analyze 1000 screens without a fully automated image analysis system and even then, in the example given at point Y, many metallographic specimens would be needed to give a total area of 1 m^2. An assessment similar to this should be made before any quantitative image analysis is started.

There is currently no British or USA standard which sets out the approach just outlined, for LM or any other type of microscopy. Inclusion levels in steels are commonly measured using charts which are inadequate for low volume fractions. Van der Voort [6] has used image analysis to measure extremely low inclusion contents. He used, as an example, the assessment of a population of inclusions with a mean size of about 1 μm and a volume fraction of about 10^{-4}. From Fig. 1 it can be seen that about 100 screens would need to be analyzed, and Van der Voort

commented that he used '5 to 500 fields' to achieve his quoted precision of ±60%. Problems limiting accuracy included the failure to measure small inclusions (below the resolving power of the microscope) and the presence of cracks or holes at the particle/matrix interfaces, which may or may not have been due to specimen preparation.

Details of the number of features measured are often omitted from published work and it is therefore difficult to assess the extent to which proper procedures are used. It is the opinion of this author, however, that many apparently quantitative measurements are based on inadequate numbers of data points. As an example of the disparity between the numbers which different authors feel are appropriate (without overt justification in either case) we can cite two sets of authors who have published detailed studies of the spatial distribution of fibres in unidirectional composites. Green and Guild [7] specified that at least 2000 fiber measurements were sorted into 30 histogram bins, while Pyrz [8] drew similar types of conclusions from only about 300 fibers sorted into an unspecified number of bins. Neither author cites a rationale for the widely differing size of data set.

5.4.2 Image Projection

Three distinct methods of sampling are implicitly used in the collection of both images and analytical data intended for quantitative use. We must distinguish between: (i) true 3D sampling, as typified by the confocal microscopies or by SIMS (relatively rare); (ii) genuine 2D sampling, for example, of a surface by STM or of an opaque section by LM; and (iii) projected 2D sampling, for example in the TEM or SEM. We will call this $2\frac{1}{2}$D sampling. This is actually the most commonly used sampling method, in which a finite thickness of specimen is imaged or analyzed as if it were projected on a two-dimensional plane.

The well-established discipline of stereology has developed many algorithms to help with the interpretation of three-dimensional structure from two-dimensional images, either of sections or projections. There are a number of good texts on this topic [9–11]. However, while the mechanics of transforming shapes viewed in 2D to the original 3D morphology is well understood (if not always practiced), the statistical significance of such observations is not often considered with the same rigor.

The application of stereology to metallographic sections, for example, to measure grain size and shape, is well established [10, 12]. It has been pointed out, however, that the obvious choice of three orthogonal sections is not always the best way to collect data about the distribution of features, particularly in anisotropic materials. Gokhale et al. [13], for example, have shown that it is more efficient to use three sections, each parallel to the major direction of anisotropy but mutually at angles of 120°, in order to determine surface area per unit volume for fiber composites.

5.5 Data Recording

5.5.1 Light Microscopy

The light microscope (LM), used in reflection mode as is normal for most materials science studies, samples a truly 2D section

of the specimen. The Chone analysis described above and in Fig. 1 is, therefore, directly applicable and standard stereology techniques [9] can easily be applied. Specimen preparation is not often a significant problem since large areas suitable for viewing can usually be prepared.

5.5.2 Scanning Electron Microscopy

The use of scanning electron microscopy (SEM) is relatively straightforward, although it must be recognized that, in secondary electron imaging mode, the SEM involves a $2\frac{1}{2}$D projection since it actually samples a surface layer of finite thickness. The volume sampled is relatively small for imaging purposes when secondary electrons only penetrate a few nanometers but is considerably greater during X-ray analysis. An EDX analysis of a solid sample is usually an average over a depth of at least 1 μm. When used in backscattered mode on polished sections the SEM appears to approach the conditions of use of the light microscope but it is in fact sampling over a projected thickness of up to 5 μm. As long as this limitation is borne in mind the SEM can provide an extension of LM techniques to smaller feature sizes. As with LM, the preparation of sufficiently large areas for proper sampling of microstructure is not usually a problem.

5.5.3 Transmission Electron Microscopy

The problem of reliable sampling on a microstructural scale is a major stumbling block to quantitative use of transmission electron microscopy (TEM). Thin areas are difficult to achieve, are rarely extensive, and may represent special regions of the specimen (e.g., those which thin fastest). There has been very little progress made towards standardization of specimen preparation techniques [4, 14]. TEM, in both imaging and analysis modes, is essentially a projection sampling ($2\frac{1}{2}$D) technique and in order to convert the 2D measurements to any sort of 3D parameter (e.g., number of features per unit volume) the section thickness must be known. This obviously raises the problem of measuring thickness. There are many ways of doing this and their relative merits have been discussed by Budd and Goodhew [15], for example. A further equally important consideration has received much less attention, namely, the thickness of 'thin films' in the materials sciences is rarely constant even over the whole area of a single micrograph. We should, therefore, measure the thickness at a large number of positions in each micrograph. This is rarely considered feasible and therefore places a limit on the accuracy of much so-called quantitative TEM data.

A further issue particular to TEM is the visibility of the features to be measured. Image contrast frequently depends on diffraction conditions and therefore it is possible, indeed common, that features of interest (dislocations, stacking faults, precipitates) are not visible in a particular micrograph. It may indeed not be practicable to find imaging conditions which reveal all the features of interest in the same micrograph. There are thus three different classes of feature which may escape detection in the TEM: (i) those below the resolution limit, (ii) those

which are out of contrast, and (iii) those which are too large to be retained in the section. These limitations must be taken into account in reporting quantitative conclusions. Although in principle each of these three factors affects the results of any technique, it is in the TEM that (ii) and (iii) are particularly important.

There are still many reasons for wishing to use TEM for quantitative hyleography. Improved resolution is in principle combined with the ability to identify phases by electron diffraction and/or X-ray analysis (EDX) or energy loss spectrometry (EELS). These techniques each need calibration, which is not usually a routine matter for most electron microscopists, and may indeed pose problems which have not yet been addressed in electron metallography. Among the items which must be calibrated are magnification and diffraction camera length. Standards in this area would require not only the specification of standard objects for magnification calibration in all magnification ranges from 100× to 1 000 000× but recommendations for frequent recalibration because of possible variations in lens currents and temperature.

Analytical techniques such as EDX and EELS are much more difficult to calibrate, since chemically homogeneous standards are not available for many elements. It is relatively easy to calibrate the detection efficiency of the analyzer but much more difficult to calibrate and standardize the computer correction routines which are used to convert raw data into composition. Attention should of course be paid to the numbers of counts in each peak or edge. This is of particular concern in TEM and STEM analyses of small particles in thin foils. Devenish et al. [16] have demonstrated the errors inherent in analyses which rely on less than 10 000 counts from the peak of interest.

5.5.4 Electron Microprobe Analysis

Analysis by electron microprobe analysis (EPMA) is a well-developed technology and some thought has been given to accuracy and precision. The items which require attention are the specification of standards and operating procedures and the calibration of routines which are used to correct for atomic number, absorption and fluorescence effects (ZAF corrections).

As with all the techniques discussed in this article there is a problem in the determination of phases present in low volume fractions. A particular aspect of this problem relates to finding a region of potential interest. It is extremely time-consuming, in EPMA and also other surface-sensitive analysis techniques such as scanning Auger microscopy (SAM), to locate small regions of interesting composition. The problem is associated with the contrast exhibited by these regions. Most imaging techniques (light microscopy, SEM, etc.) produce an image virtually instantaneously and if features of interest give rise to distinct contrast (e.g., appear dark) it is easy to detect them by eye or by image analyzer. However, phases of interesting composition may not be easy to reveal in the image of an SEM, EPMA or SAM instrument. They must then be sought on the basis of their composition. Compositional mapping is generally very time consuming, typically taking a factor 10^4 or 10^5 longer than conventional imaging of the same region. Search

algorithms are therefore needed to speed up identification of interesting regions, before quantitative assessment can start. Baker and Castle [17] have made a start on this, in the context of surface analysis, but much more work is still needed.

5.5.5 Scanning Probe Techniques

Scanning probe techniques (SPM) such as STM and AFM are usually genuinely two-dimensional, in that they sample a true surface which is only about one atom in thickness. The projection problem is, therefore, scarcely applicable, but the statistical sampling problem remains.

At present SPM techniques are still mainly being used in a qualitative way, in the sense that observations are being made of single examples of behavior and little attempt is being made to measure, for instance, the area density of adatoms or surface clusters. When the techniques mature they are certain to be used in a more quantitative manner and analyses similar to the Chone analysis described above will need to be made. One example will serve to illustrate the approach:

Assume that an STM (it could just as well be an AFM) is being used to image atomic-scale details. A typical field of the surface, viewed at a few million times magnification, is unlikely to contain more than 10000 atoms. Suppose that we need to measure the concentration (number per unit area) of surface features at the 10% level (i.e., 10% of the surface layer of atoms are involved in the features of interest). To measure this concentration to an accuracy of 20% (i.e., to place it in the range $10 \pm 2\%$) requires the measurement of a number of micrographs which depends on the size of the feature. When many micrographs are used, edge effects need to be taken account of since a feature must fall entirely within a field to be easily automatically recognized. An unsophisticated analysis, using the above criteria with 10000 atoms per micrograph, shows that mutliple micrographs will be needed if clusters containing more than 10 atoms are to be measured. The results are summarized in Table 1.

5.5.6 Other Techniques

There are many other techniques to which a proper analysis of sampling and statistics should be applied. A number of these (e.g., SIMS, RBS and XPS) generate large

Table 1. The number of micrographs needed to estimate the concentration of surface features to an accuracy of ±20% using an atomic-resolution surface imaging technique

Description of feature	Approximate number of atoms per feature	Average number of features per micrograph	Number of micrographs needed
single atom	1	1000	1
2-atom cluster	2	492	1
10-atom cluster	10	92	1
7×7 reconstruction on Si	100	6	4
6 nm × 6 nm cluster	400	0.6	40

numbers of counts, giving the capability for a statistically precise analysis averaged over quite a large area. However, because they offer good resolution in only one direction (usually depth), they are not really mainstream hyleographic techniques since they offer poor discrimination of microstructural features.

There is at least one technique, atom probe field ion microscopy (APFIM) which offers excellent spatial resolution of microstructure in three dimensions. However, the problems of proper unbiased sampling are extreme, because of the small volume which can be accessed and the small numbers of counts which contribute to any analysis. This problem is discussed further in the following section.

5.6 Ultimate Sensitivity and its Significance

There is an understandable drive in the hyleographic community to improve the ultimate sensitivity of each technique. For imaging techniques this is expressed as an improvement in resolution while for analytical techniques it is a reduction in the minimum detectable amount. There is, however, a price to be paid for sensitivity and this should be appreciated.

In an imaging instrument an improvement in resolution is usually accompanied by an increase in the commonly-used magnification, and therefore a decrease in the amount of material sampled in each micrograph. This is fairly obvious, and easy to assess, for 2D sampling techniques such as LM and SPM and for $2\frac{1}{2}$D techniques with a small sampling depth such as SEM. In these cases it is straightforward to increase the area sampled by taking more micrographs. However, for the TEM, the situation is made worse by the fact that higher resolution is often accompanied by the need for reduced specimen thickness. This both reduces the sampled volume and makes thickness measurements relatively less precise. It is then rarely possible to take enough micrographs to properly satisfy a Chone-type sampling analysis.

Let us consider, as an example, the detection limit in analytical STEM. It is easy to show that, with a modern dedicated STEM and large collection angle X-ray detectors [18], it should be possible to detect five atoms or fewer. However, these would be located in a column containing 40 000 matrix atoms and would represent a local concentration of about 0.01%. If we attempt to take this beyond a demonstration of sensitivity and draw quantitative conclusions about the concentration and/or distribution of these atoms we need to consider whether the presence of five atoms in this small volume has any wider significance, and whether the difference between four, five and six atoms in this volume can be taken seriously. The two key points are (i) an error of plus or minus one atom represents a $\pm 20\%$ uncertainty in the measured concentration, and (ii) the extrapolation of this measurement to a conclusion about the overall concentration in a larger part of the specimen would demand an almost incredible assumption of homogeneity. For these reasons analyses at the ultimate sensitivity should remain essentially qualitative.

The statistics involved in APFIM analysis are subject to a similar caution. Numakura [19] has pointed out recently that it is unsound to draw quantitative

conclusions on the compositions of small particles when the number of ions analyzed per particle can be as small as seven. In the case of APFIM it is particularly difficult to collect statistically meaningful data from a large enough population of particles so the conclusion drawn concerning very small particles and clusters (which, admittedly, can be analyzed in no other instrument) should remain largely qualitative.

There is, however, a more positive view of analytical sensitivity. Kisielowski et al. [20] have commented that, in a study of diffusion, their rather large analyzed volume of 10^5 $(nm)^3$ might have contained about 10^8 atoms, but on average it contained only a single vacancy. The analysis was therefore sensitive to the *effect* of a single vacancy. This is an interesting viewpoint, which is not often articulated.

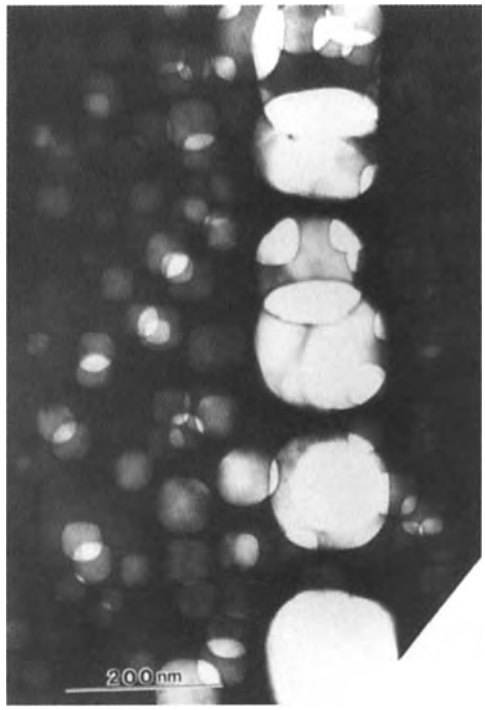


Figure 2. Gas bubbles in a cross-sectional specimen of helium-implanted nickel.

5.7 Case Studies

5.7.1 Cavity Size Distributions in Irradiated Metals

During studies of the evolution of populations of bubbles and voids in irradiated materials it is often desirable to measure the cavity number and volume from TEM micrographs (e.g., [21]).

As Fig. 2 shows, these features can range in size from 1 nm to several hundred nm and the distribution of cavities is rarely uniform within the foil. The problem here is to find enough volume to sample. As a result of ion implantation the size and density of bubbles varies drastically with depth below the implanted surface. In a plan-view specimen the bubble density is thus not likely to be constant throughout the foil thickness, making quantitative measurement almost impossible without the use of very time-consuming stereo techniques. In a cross-sectional specimen such as Fig. 2 a second problem arises: there is then not a great volume available in which each typical population can be imaged. Measurements on uniform populations of bubbles have usually been made with at least 1000 bubbles per population. However, in the specimen shown in Fig. 2 this is impracticable, particularly for the larger bubble sizes [22]. Even at small bubble sizes image overlap is a problem. Quantitative measurements are inevitably imprecise and in this particular case only qualitative conclusions were drawn.

5.7.2 Dislocation Spacing Distributions in Semiconductor Interfaces

In assessing the relaxation at pseudomorphic epitaxial interfaces it is frequently necessary to measure dislocation spacings in extensive arrays (e.g., [23]). The specimen is essentially a bicrystal and the two orthogonal sets of dislocations (e.g., Fig. 3) may have different spacings. The spacing is locally irregular and in this case it is necessary to image two orthogonal long thin strips of the specimen, in order to be able to count the number of dislocations in each direction. It is in practice unusual to find that a specimen has thinned so that it is electron-transparent for long distances in two orthogonal directions and statistical compromises generally have to be made.

The dislocation spacing is random and the problem is very similar to that associated with counting statistics in EDX or EELS analysis. The noise to signal ratio (equivalent here to the error:measurement ratio) is given by $1:\sqrt{N}$ where N is the number of dislocations whose spacing is measured. To obtain an accuracy of 10% we therefore need to measure the average spacing of 100 dislocations. If they are about 0.5 μm apart (as in Fig. 3) then this requires a thin region 50 μm long in each of the two orthogonal directions.

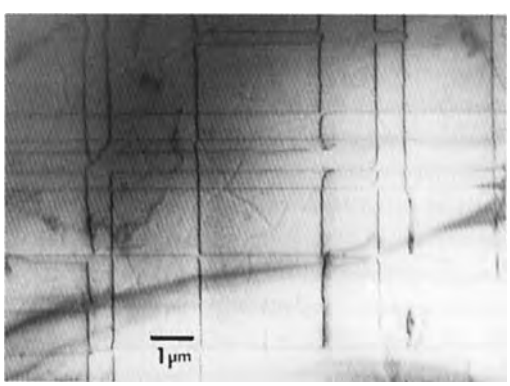


Figure 3. A network of interfacial dislocations between InGaAs and GaAs.

In addition to this problem there is the problem associated with sampling. The TEM specimen samples much less than $1\,mm^2$ of the total wafer area which would typically be about $5000\,mm^2$. It is well known that grown-in dislocation density can vary widely across a wafer, so that array spacings are likely to vary equivalently. In what appears to be the only detailed study of this variation in the TEM, Yeo [24] found a scatter of more than a factor three among 19 specimens from the same wafer. Clearly, it is not sufficient that the statistics of the measurement on a single specimen are good.

In a more recent study of the distribution of dislocation spacings in InGaAs/GaAs interfaces Macpherson et al. [25] have had to measure more than 1000 spacings to characterize a single set of dislocations. In matching their experimental data to theoretical predictions they also had to take account, quantitatively, of the finite image width of dislocations. When measuring spacings of parallel dislocation lines which are not all in the same plane, data concerning small spacings are lost because the images of dislocations whose projected images are close together are seen as one line. In this study a quantitative estimate of image width in known diffraction conditions was incorporated into the model of the predicted distribution and a good match with the experimental spacings was obtained. This level of sophistication is, not surprisingly, rarely

reported because it is very time-consuming. Nonetheless, it should become increasingly common.

5.8 References

[1] P. J. Goodhew, *Microscopy and Analysis*, Jan. **1993,** 11.
[2] E. R. Daniel, *Met. Mater.* **1991,** *7*, 522.
[3] M. J. C. de Jong, *Inst. Phys. Conf. Series* **1988,** *93*, 141.
[4] P. J. Goodhew in *Materials Metrology and Standards for Engineering Design* (Ed.: B. F. Dyson), Elsevier Science **1992,** p. 271.
[5] J. Chone, *International Symp. Quantitative Metallography*, Florence (Ed.: Association Italiana di Metallurgia) **1978,** p. 209.
[6] G. F. Vander Voort in *Effect of Steel Manufacturing Processes on the Quality of Bearing Steels*, ASTM STP 987 (Ed.: J. J. C. Hoo), **1988,** p. 226.
[7] D. Green, F. J. Guild, *Composites* **1991,** *22*, 239.
[8] R. Pyrz, *Compos. Sci. Technol.* **1994,** *50*, 197.
[9] E. E. Underwood, *Quantitative Stereology*, Addison–Wesley, Reading, MA **1970.**
[10] J. C. Russ, *Practical Stereology*, Plenum Press, London **1986.**
[11] R. T. Dehoff, F. N. Rhines, *Quantitative Microscopy*, McGraw–Hill, New York **1968.**
[12] J. R. Blank, T. Gladman, in *Tools and Techniques in Physical Metallurgy* (Ed.: F. Weinberg) **1970,** Chapter 4, p. 265.
[13] A. Gokhale, W. J. Drury, B. Whited, *Mater. Charact.* **1993,** *31*, 11.
[14] P. J. Goodhew, *Thin Foil Preparation for Electron Microscopy*, Elsevier, Oxford **1985.**
[15] P. M. Budd, P. J. Goodhew, *Light Element Analysis in the Transmission Electron Microscope*, Oxford Science Publications, Oxford, UK **1988.**
[16] R. W. Devenish, G. J. Tatlock, B. L. Eyre, Inst. Phys. Conf. Ser. (UK) No. 68, **1983,** 289.
[17] M. A. Baker, J. E. Castle, *Proc. EUREM 88, Inst. Physics Conf. Ser.* (UK) **1988,** *93*, 269.
[18] G. Cliff, R. W. Devenish, P. J. Goodhew, R. J. Keyse, G. W. Lorimer, *Proc. 13th Int. Cong. EM, Paris* **1994,** *1*, 719.
[19] H. Numakura, *Scr. Metall. Mater.* **1994,** *30*, 661.
[20] C. Kisielowski, P. Schwander, Y. Kim, J.-L. Rouviere, A. Ourmazd, *Phys. Status Solidi a* **1993,** *137*, 557.
[21] P. J. Goodhew, S. K. Tyler, *Proc. R. Soc. Lond. A* **1981,** *377*, 151.
[22] N. Marochov, L. J. Perryman, P. J. Goodhew, *J. Nucl. Mater.* **1987,** *149*, 296.
[23] P. J. Goodhew, *J. Phys. & Chem. Solids* **1994,** *55*, 1107.
[24] Y. K. Yeo, MSc thesis, University of Surrey, UK **1990.**
[25] G. MacPherson, R. Beanland, P. J. Goodhew, *Philos. Mag.* **1995,** *72*, 1531.

Handbook of Microscopy
S. Amelinckx, D. van Dyck, J. van Landuyt, G. van Tendeloo

General Reading

Chapter I.1

P. B. Hirsch, A. Howie, R. B. Nicholson, D. W. Pashley, M. J. Whelan, *Electron Microscopy of thin Crystals*, Butterworth, London, **1965**.
S. Amelinckx, R. Gevers, G. Remaut, J. van Landuyt (Eds.), *Modern Diffraction and Imaging Techniques in Material Science*, North-Holland, Amsterdam, **1970**.
G. Thomas, M. J. Goringe, *Transmission Electron Microscopy of Materials*, Wiley, New York, **1979**.
L. Reimer, *Transmission Electron Microscopy*, Springer Verlag, Berlin, **1984**.
J. C. H. Spence, *Experimental High-Resolution Electron Microscopy*, Oxford University Press, Oxford, **1986**.
J. M. Cowley, *Diffraction Physics*, North-Holland, Amsterdam, **1990**.
P. Buseck, J. Cowley, L. Eyring, *High Resolution Transmission Electron Microscopy and associated Techniques*, Oxford Science Publications, Oxford, **1992**.
J. M. Cowley (Ed.), *Electron Diffraction Techniques*, Volumes 1 and 2, Oxford Science Publications, Oxford, **1992**.
G. R. Anstis and J. L. Hutchinson, in: *Dislocations in Solids*, ed. F. R. N. Nabarro, Vol. 9, **1992**, p. 1-56.
M. H. Loretto, *Electron beam Analysis of Materials*, Chapman & Hall, London, **1994**.
J. W. Christian, *The Theory of Transformations in Metals and Alloys*, Pergamon Press, **1965**.
A. Khatchaturyan, *Theory of Structural Transformations in solids*, John Wiley & Sons, **1983**.
R. W. Cahn, *Physical Metallurgy*, North-Holland, **1965**.
D. Porter & K. Easterling, *Phase Transformations in Metals and Alloys*, VNR, **1988**.
P. Haasen (Ed.), *Phase Transformations in Materials*, VCH, **1991**.
Z. Nishiyama, *Martensitic Transformations*, Academic Press, **1978**.
G. Stocks and A. Gonis (Ed.), *Alloy Phase stability*, NATO ASI series, Vol. 163, **1987**.
A. Gonis, P. Turchi and J. Kudrnovsky (Eds.), *Stability of Materials*, NATO ASI series, Vol. 355, **1994**.
R. Krishnan, S. Banerjee and P. Mukhopadhay (Eds.), *Materials Science Forum: Phase stability and Phase Transformations*, Trans. Tech. Publ., **1985**.

Chapter I.2

P. R. Buseck (Ed.), *Minerals and Reactions at the Atomic Scale, Rev. Mineral. 27*, Mineralogical Society of America, Washington, DC, **1992**.
L. M. Coyne, S. W. S. McKeever, D. F. Blake (Eds.), *Spectroscopic Characterization of Minerals and Their Surfaces*, American Chemical Society, Washington, DC, **1990**.
W. A. Deer, R. A. Howie, J. Zussman, *An Introduction to the Rock-forming Minerals*, 2nd ed., Longman Scientific and Technical, Harlow, **1992**.
J. F. Kerridge, M. S. Matthews (Eds.), *Meteorites and the Early Solar System*, Univ. of Arizona Press, Tucson, **1988**.
C. Klein, C. S. Hurlbut, Jr., *Manual of Mineralogy (after J. D. Dana)*, 20th ed., Wiley, New York, **1985**.
H. A. Lowenstam, S. Weiner, *On Biomineralisation*, Oxford University Press, New York, **1989**.
A. S. Marfunin (Ed.), *Advanced Mineralogy*, Vols. 1 & 2 (Volume titles are *1. Composition, Structure and Properties of Mineral Matter; 2. Methods and Instrumentations)*, Springer Verlag, Berlin, **1994**.
A. C. McLaren, *Transmission Electron Microscopy of Minerals and Rocks*, Cambridge University Press, Cambridge, **1991**.
W. D. Nesse, *Introduction to Optical Mineralogy*, 2nd ed., Oxford University Press, Oxford, **1991**.
P. J. Potts, J. W. F. Bowles, S. J. B. Reed, M. R. Cave (Eds.), *Microprobe Techniques in the Geosciences*, (Min. Soc. Series, UK), Chapman and Hall, London, **1995**.
A. Putnis, *Introduction to Mineral Sciences*, Cambridge University Press, Cambridge, **1992**.

Chapter I.3

S. M. Sze, *VLSI-Technology*, McGraw-Hill, New York, **1988**.
S. K. Ghandhi, *VLSI Fabrication Principles*, John Wiley, New York, **1994**.
F. Shimura, *Semiconductor Crystal Technology*, Academic Press, London, **1989**.

S. Mahajan (Ed.), *Handbook on Semiconductors, Materials Properties and Preparation*, Vols. 3A, B, North-Holland, Amsterdam, **1994**.
T. Kamins, *Polycrystalline Silicon for Integrated Circuit Applications*, Kluwer, Boston, **1988**.
S. P. Murarka, *Metallization – Theory and Practice for VLSI and ULSI*, Butterworth-Heinemann, London, **1993**.
P. H. Holloway, G. E. McGuire (Eds.), *Handbook on Compound Semiconductors*, Noyes, Park Ridge, NJ, **1996**.
S. M. Sze, *High-Speed Semiconductor Devices*, J. Wiley, New York, **1990**.
E. Fuchs, H. Oppolzer, H. Rehme, *Particle Beam Microanalysis*, VCH, Weinheim, **1990**.
H. Cerva, H. Oppolzer, *Microstructure and Interfaces of Polysilicon in Integrated Circuits*, Springer Proc. Phys. 35, **1989**, p. 354.
J. Ehrstein, R. Mathur, G. McGuire (Eds.), *Proc. 3rd Int. Workshop on Measurement and Characterization of Ultra-Shallow Doping Profiles in Semiconductors*, **1995**.
B. O. Kolbesen, H. P. Strunk, *Analysis, Electrical Effects and Prevention of Process-Induced Defects in Silicon Integrated Circuits, VLSI Electronics Microstructure Science* (Ed.: G. Einspruch), Academic Press, New York, Vol. 12, **1985**, p. 143.
H. Cerva, H. Oppolzer, *Characterization of Heteroepitaxial Compound Semiconductor Layers and Superlattices using TEM*, Prog. Crystal Growth Charact., 20, 231, **1990**.
Proc. Microsc. Semicond. Mater., Inst. Phys. Conf. Ser., **1981-1995**.
Proc. Int. Symp. for Testing and Failure Analysis (ISTFA/90-95), **1990-1995**.
Proc. Europ. Symp. Reliability of Electronic Devices, Failure Physics and Analysis (ESREF 90-95), **1990-1995**.

Chapter I.4

S. M. Sze, in: *Physics of Semiconductor Devices*, 2nd Ed., Wiley & Sons, New York, Chapter 12, p. 681, **1981**.
C. R. M. Grovenor, in: *Microelectronic Materials* (Ed.: B. Cantor), Adam Hilger Press, Bristol, **1989**.
J. Wilson, J. F. B. Hawkes, in: *Optoelectronics: An introduction* (Ed.: P. J. Dean), Prentice-Hall, Englewood Cliffs, NJ, **1983**.
J. W. Edington, in: *Electron Diffraction in the Electron Microscope*, Philips Technical Library, Eindhoven, **1975**.
G. L. Bir, G. E. Pikus, in: *Symmetry and Strain-Induced Effects in Semiconductors*, Wiley & Sons, New York, **1974**.
J. B. Clegg, in: *Growth and Characterization of Semiconductors* (Ed.: R. A. Stradling and P. C. Klipstein), Adam Higler, Bristol, **1990**, p. 87.
D. B. Holt, in: *Growth and Characterization of Semiconductors* (Ed.: R. A. Strading, P. C. Klipstein, Adam Higler, Bristol, **1991**, p. 56.
E. Rosencher, in: *Silicon Molecular Beam Epitaxy*, (Ed.: E. Kasper, J. C. Bean), Vol.1, p. 161, CRC Press, Boca Raton, **1988**.
S. Luryi and S. M. Sze, in: *Silicon Molecular Beam Epitaxy*, (Ed.: E. Kasper, J. C. Bean), Vol. 1, **1988**, p. 181.
J. C. Bean, in: *Silicon Molecular Beam Epitaxy*, (Ed.: E. Kasper, J. C. Bean), Vol. 2, CRC Press, Boca Raton, **1988**, p. 65.

Chapter I.5

E. K. H. Salje, *Phase Transitions in Ferroelastic and Co-elastic Crystals*, Student ed. Cambridge University Press, Cambridge, **1993**.

Chapter I.7

J. E. Arem, *Color Encyclopedia of Gemstones*, Van Nostrand Reinhold, New York, **1977**.
E. Bruton, *Diamonds*, 2nd ed., N.A.G. Press, London, **1981**.
E. J. Gübelin, *Internal World of Gemstones*, 3rd ed., ABC Edition, Zürich, **1983**.
E. J. Gübelin, J. I. Koivula, *Photoatlas of Inclusions in Gemstones*, 2nd ed., ABC Edition, Zürich, **1992**.
G. Lenzen, *Edelsteinbestimmung mit gemmologischen Geräten*, Verlagsbuchhandlung Elisabeth Lenzen, Kirschweiler, **1984**.
R. Webster, *Gems: their Sources, Descriptions and Identification*, 4th ed. (revised by B. Anderson, Butterworths, London, **1983**.

Chapter I.8

M. Cyrot and D. Pavuna, *Introduction to Superconductivity and High Tc Materials*, World Scientific, Singapore, **1992**.
R. Kossowsky, S. Methfessel and D. Wohlleben (Eds.), *Physics and Materials Science of High*

Temperature Superconductors, Kluwer Academic Publishers, Dordrecht, **1990**.
E. Kaldis (Ed.), *Materials and Chrystallographic Aspects of HTc-Superconductivity*, Kluwer Academic Publishers, Dordrecht, **1994**.
S. E. Babcock and K. L. Merkle (Eds.), in: *Special Issue on Interfaces in High Tc Superconductors*, Interface Science, Volume 1, nr 4, **1993**.
Materials Characterisation: Special feature, *Superconducting Materials*, volume 27, **1991**.
F. C. Matacotta, G. Ottaviani (Eds.), *Science and Technology of Thin Films*, World Scientific, Singapore, **1995**.
Special issue on imaging of High Tc Superconducting Materials. *Journal of Electron Microscopy Technique 8*, **1988**.

Chapter I.9

P. R. Buseck, J. M. Cowley and L. Eyring (EDS), *High Resolution Transmission Electron Microscopy and Associated Technique*,- Particularly chapter 14 by A. Howie, Chapter 12 by O. L. Krivanek and Chapters 1-3 by J. M. Cowley, New York, Oxford, Oxford University Press, **1992**.
J. C. H. Spence, *Experimental High-Resolution Electron Microscopy*, Oxford, Clarendon-Press, **1981**.
P. H. Gaskell, in: *Electron Beam Imaging of non-Crystalline Material*, (Ed. K. M. Knowles) IOP Short Meetings Series (11), Bristol, Institute of Physics, 47, **1988**.
A. Howie, *J. Non-Cryst. Solids*, *31*, **1978**, 41.
P. H. Gaskell, in: *Materials Science and Technology A comprehensive review*, (Eds.: R. W. Cahn, P. Haasen and E. J. Kramer) VCH Weinheim, Germany, Vol. 9 (Ed. J. Zarzycki) **1991**, 175.
J. Zarzycki, *Glasses and the vitreous state*, Cambridge Solid State Science Series, Cambridge University Press, Cambridge, **1982**.
A. I. Goldman and K. Widom, *Annu. Rev. Phys. Chem.*, *42*, **1991**, 685.
C. Janot, *Quasicrystal: A Primer*, Oxford University Press, Oxford, **1992**.
D. Gratias, *Contemp. Phys.*, *28*, **1987**, 219.
P. Guyot, P. Kramer and M. de Boisseau, *Reports on Progress in Physics*, *54*, **1991**, 1373.
M. V. Jaric (Ed.), *Aperiodicity and Order*, Vol. 1, *Introduction to Quasicrystals*, Academic Press, Boston, **1988**; Vol. 2, *Introduction to the Mathematics of Quasicrystals*, Academic Press, Boston, **1989**; M. V. Jaric and D. Gratias (Eds.), Vol.3, *Extended Icosahedral Structures*, Academic Press, Boston, **1989**.
K. F. Kelton, in: *Intermetallics Compounds*, (Eds. J. H. Westbrook, R. L. Fleischer), Wiley, New York, Chap. 20, **1994**.
K. H. Kuo, *Mater. Chem. Phys.*, *39*, **1994**, 1.
S. Ranganathan and K. Chattopadhyay, *Annu. Rev. Mater. Sci.*, *21*, **1991**, 437.
M. J. Yacaman, M. Torres (Eds.), *Crystal-Quasicrystal Transition*, North-Holland, Amsterdam, **1993**.
A. Yamamoto, *Acta Cryst. A*, *52*, **1996**, 509.

Chapter I.10

B. D. Ratner (Ed.), *Surface Characterization of Biomaterials, Progress in Biomedical Engineering*, Amsterdam, Elsevier, **1988**.
D. F. Williams (Ed.), *Concise Encyclopedia of Medical and Dental Materials*, Oxford, Pergamon Press, **1990**.
D. F. Williams (Ed.), *Medical and Dental Materials, Science and Technology*, Vol. 14, Weinheim, VCH, **1992.**
K. Yasuda, *Age-Hardening and Related Phase Transformation in Dental Gold Alloys,* Gold Bull., 20, **1987**, 90-103.
K. Yasuda and K. Hisatsune, *Microstructure and Phase Transformation in Dental Gold Alloys, Determination of a Coherent Phase Diagram*,- Gold Bull., *26*, **1993**, 50-66.

Chapter I.13

J. Spence, *Experimental High Resolution Electron Microscopy*, Oxford University Press, New York, Oxford, **1988**.
J. Cowley, *Diffraction Physics*, North-Holland, Amsterdam, Oxford, NY, Tokyo, **1986**.
P. W. Atkins, *Molecular Quantum Mechanics*, Oxford University Press, Oxford, NY, **1994**.
D. C. Champeney, *Fourier Transforms and their Physical Applications*, Academic Press, London, **1973**.
N. E: Cusack, *The Physics of Structurally disordered Matter*, Adam Hilger, Bristol, Philadelphia, **1987**.
C. Giacovazzo, H. L. Monaco, D. Viterbo, F. Scodari, G. Gilli, G. Zanotti, M. Catti (Ed. C. Giacovazzo), *Fundamentals of Crystallography*, International Union of Crystallography, Oxford University Press, Oxford, **1992**.

D. Dorset, *Structural Electron Crystallography*, Plenum Press, New York, London, **1995**.
B. Buck, V. Macanlay (Ed.), *Maximum Entropy in Action*, Oxford Science Publication, Clarendon Press, Oxford, **1991**.
P. G. de Gennes, *The Physics of liquid crystals* Clarendon Press, Oxford, **1974**.
Gordon and Breach, *Ferroelectric liquid Crystals. Ferroelectricity and related phenomena*, Vol. 7, Science Publishers, Philadelphia, **1991**.
D. S. Chemla, J. Zyss (Ed.), *Non-linear optical properties of organic molecules and Crystals*, Vol. 1, Academic Press, Inc. Orlando, San Diego, NY., Boston, London, Tokyo, Toronto, **1987**.
D. L. Misell, *Image Analysis, Enhancement and Interpretation*, North-Holland, Publ. Co., Amsterdam, NY., Oxford, **1978**

Chapter I.14

Hj. Matzke, *Science of Advanced LMFBR Fuels*, North-Holland, Amsterdam, **1986**.
B. Frost, *Nuclear Fuel Elements*, Oxford, Pergamon, **1982**.
J. Gittus, *Irradiation Effects in Crystalline Solids*, Applied Science Publ., London, **1978**.
W. Lutze, R. C. Ewing (Eds.), *Radioactive Waste Forms for the Future*, North-Holland, Amsterdam, **1988**.
Proc. of the Annual Symposia on the Scientific Basis for Nuclear Waste Management, since **1986**.
T. A. Abrajano jr., L. H. Johnson (Eds.), *Mat. Res. Soc. Proc.*, Vol. 212, **1991**.

Chapter I.15

C. Kittel, *Rev. Mod. Phys., 21*, 541, **1949**.
C. Kittel, J. K. Galt, *Solide State Phys., 3*, 437, **1956**.
D. J. Craik, R. S. Tebble, *Ferromagnetism and Ferromagnetic Domains*, North-Holland, Amsterdam, **1965**.
R. Carey, E. D. Isaac, *Magnetic Domains and Techniques for Their Observation,* Academic Press, London, **1966**.
M. Rosenberg, C. Tanasoiu, Magnetic Domains, in: *Magnetic Oxides,* (Ed. D. J. Craik), J. Wiley, London, **1972**.
D. J. Craik, The observation of magnetic domains, in: *Methods of Experimental Physics* (Ed. R. V. Coleman), Academic Press, New York, **1974**.
R. P. Ferrier, Imaging methods for the study of micromagnetic structure, in: *Noise in Digital Magnetic Recording*, (Ed. T. C. Arnoldussen, L. L. Nunneley) World Scientific, Singapore, **1992**.
A. Hubert, R. Schäfer, *Magnetic Domains*, Springer Verlag, Berlin, Heidelberg, New York, in preparation, planned for **1998**.

Chapter II.3

D. A. Kay (Ed.), *Techniques for Electron Microscopy*, Blackwells, Oxford, **1965**.
P. B. Hirsch, R. B. Nicholson, A. Howie, D. W. Pashley, M. J. Whelan, *Electron Microscopy of Thin Crystals*, Butterworths, London, **1965**.
G. Schimmel, W. Vogel (Ed.), *Methodensammlung der Elektronenmikroskopie*, Wissenschaftliche, Stuttgart, **1973-1979**.
A. M. Glauert (Ed.), *Practical Methods in Electron Microscopy*, North-Holland, London, **1972**.
T. Mulvay and R. K. Webster (Ed.), *Modern Physical Techniques in Materials Technology*, Oxford University Press, Oxford, **1974**.
K. C. Thomson-Russel, J. W. Edington, *Practical Electron Microscopy in Materials Science 5: Electron Microscopy Preparation Techniques in Material Science*, N. V. Philips' Gloeilampfabriken, Eindhoven, **1977**.
P. J. Goodhew, *Specimen Preparation for Transmission Electron Microscopy, Microscopy Handbooks, Royal Microscopical Society*, Oxford University Press, **1984**.
R. Anderson (Ed.), *Workshop on Specimen Preparation for TEM of Materials II. MRS Symposium Proceedings*, Vol. 199, MRS, Pittsburgh, **1990**.
A. W. Robards, A. J. Wilson (Ed.), *Procedures in Electron Microscopy*, Wiley, Chichester, **1993**.

Chapter II.4

J. Buffle, H. P. Van Leeuwen (Eds.), *Environmental Particles*, Vol. I, Lewis, Chelsea MI, **1992**.
M. Grasserbauer, *Mikrochim*, Acta, III, 415, **1983**.

W. Jambers, L. De Bock, R. van Grieken, *Analyst, 120*, **1995**, 681.
D. Perret, G. G. Leppard, M. Müller, N. Befzile, R. De Vitre, J. Buffle, *Wat. Res., 25*, **1991**, 1333.
R. Van Grieken, C. Xhoffer, *J. Anal. Atom. Specrom.*, *7*, **1992**, 81.

Chapter II.6

P. J. Goodhew and F. J. Humphreys (Ed.), *Electron Microscopy and Analysis*, 2-1, Taylor & Francis, **1988**.
J. C. Russ, *Computer-Assisted Microscopy*, Plenum Press, **1990**.
D. B. Williams and C. B. Carter, *Transmission Electron Microscopy: A Textbook for Materials Science*, Plenum, **1996**.

Handbook of Microscopy
S. Amelinckx, D. van Dyck, J. van Landuyt, G. van Tendeloo

List of Symbols and Abbreviations

Symbol	Designation	Chapter
A	area	I.12
A	coefficient of the quadratic term of a Landau potential	I.5
a	lattice parameter	I.1
$A(u)$	aperture function	I.1
a, b, c, d, α, β, γ	lattice parameters of the low-symmetry phase	I.5
a, c	Cartesian coordinates	I.7
a, c	lattice parameters	I.11
a_0	lattice parameter of orthorhombic unit cell	I.8
$a_0, b_0, c_0, \alpha_0, \beta_0, \gamma_0$	lattice parameters of the high-symmetry phase	I.5
A_1	aperture function	I.13
$\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_1'$, etc.	base vectors	I.11
$a_{1,\,p}$, $a_{2,\,p}$	peraskite lattice parameter	I.8
$(A_A)_r$	reinforcement area fraction	I.12
A_A	area per unit area	I.12
a_A, a_B, a_C	length of monomer units in subchain A, B, C	I.13
$a_{C\text{-}C}$	nearest neighbor carbon–carbon distance	I.11
$a_{i,\,j}$	short-range order parameter	I.1
a_{II}	a-lattice parameter of the Ortho II structure of YBCO	I.8
a_m	mirror radius	I.12
a_m, $\mathbf{a}_m$	monoclinic lattice parameter	I.8
a_p	peraskite lattice parameter	I.8
a_T	tetragonal a parameter	I.8
A_V	area per unit volume	I.12
B	coefficient of the quartic term of a Landau potential	I.5
B	elastic constant for compression	I.13
B	magnetic flux density	I.8
$\boldsymbol{b}$	Burgers vector	I.1, I.9.2, I.11, I.14
$\boldsymbol{B}(\mathbf{r})$	magnetic flux density	I.15
$b_1, \ldots, b_6$	components of $\boldsymbol{b}$	I.9.2
$\boldsymbol{b}_i$	slip direction	I.1
c	concentration point defects	I.1
c	constant	I.13
$\boldsymbol{c}$	vector from the dislocation line to the incident beam cross-over	I.9.2
C_b	bubble concentration	I.1
CCN	cloud condensation nuclei	II.4
$\boldsymbol{c}_h$	chiral vector	I.11
C_s	spherical aberration	I.1, I.13
D	average particle diameter	II.2
D	grain size	II.3

Symbol	Designation	Chapter
D	length of obstacles	I.1
d	deformation depth of mechanical thinning	II.3
d	interplanar spacing	I.9.1
d	lattice repetition length perpendicular to a wall	I.5
d	size of phase-separated domain	I.13
d	thickness of precipitate	I.1
d	width	I.11
D'	mean length of obstacles in the glide plane	I.1
d'	mean thickness of obstacles	I.1
$D(\boldsymbol{k})$	damping function	I.13
D_1	damping envelope due to defocus spread	I.13
d_{110}	width of the twin band	I.8
D_2	damping envelope due to beam divergence	I.13
D_3	damping envelope due to film response	I.13
d_a	amorphous layer thickness	II.3
d_c	crystal thickness	I.13
$\boldsymbol{D}_{mo}$	magnetooptically induced	I.15
D_O	optical depth	I.7
D_R	real depth	I.7
d_x	thickness	I.12
E	accelerating voltage	II.3
$\boldsymbol{E}$	electric field	I.13
$\boldsymbol{E}$	electrical vector	I.15
E	energy	I.14
E	number of edges	I.11
E	potential energy; energy	I.13, I.14
$\hat{e}$	orientation of a fluctuation	I.5
$E^{\perp}$	perpendicular orthogonal 3D space	I.9.2
$E^{\parallel\parallel}$	parallel orthogonal 3D space	I.9.2
$E^{\parallel}_{max}$	maximum energy density of laser light	II.2
$E2$	accelerating beam energy	I.13
E_{Coul}	Coulomb energy	I.13
E_h	normalized structural factor	I.13
E_{hb}	hydrogen-bond energy	I.13
e_i	macroscopic strain of component i (Voigt notation)	I.5
$e_{ik}(S_i)$	macroscopic strain of component ik (matrix notation) (e.g. e_{xx}, e_{xy} ...) of the domain S_i	I.5
e_s	scalar spontaneous strain	I.5
E_{tors}	subrotation-interaction energy	I.13
E_{vdW}	van der Waals energy	I.13
E_w	energy of the wall	I.5
F	free energy density; Fourier transform	I.13
F	number of faces	I.11
F	structure factor	I.3
f	atomic scattering amplitude	I.3
f	conversion factor for grays to rems	I.14
f	nominal reinforcement volume fraction	I.12
$f(\omega)$	orientation distribution	I.13
F_c	focal point of culet	I.7

Symbol	Designation	Chapter
F_c	free energy per chain; calculated structure factor; calculated intensity	I.13
f_c	effective force for the formation of corners between domain walls	I.5
F_c^{surf}	surface free energy	I.13
$F^{elastic}$	elastic free energy	I.13
F_h	electron scattering factor	I.13
F_i	focal point of inclusion	I.7
f_i	atomic scattering factor of atom i	I.13
f_j	atomic form factor	I.1
f_j	probability of intensity I_j	I.9.1
F_N	number of hydride platelets in Zircaloy oriented parallel to a defined direction	I.14
F_o	observed structure factor; observed intensity	I.13
F_s	focal point of surface	I.7
F_t	focal point of table	I.7
f_v	volume fraction of bubbles	I.1
G	Gibbs free energy	I.13
G	shear modules	I.1
g	coefficient of the gradient energy of a Landau potential (= Ginzburg coefficient)	I.5
$\boldsymbol{g}$	diffraction vector	I.11, I.14
$\boldsymbol{g}$	reciprocal lattice vector	I.1
$\underline{\boldsymbol{g}}$	reciprocal vector in reciprocal space	I.9.2
$\overline{G}$	average shear modules	I.1
$g(2)$	pair correlation function	I.13
$\{\mathbf{H}\}$	basis set	I.13
h	homogeneity	I.12
h	Miller index	I.1
h	number of hexagons	I.11
h	reduced magnetic	I.15
$H(\boldsymbol{k})$	Fourier transform of transfer function	I.13
$\boldsymbol{H}(\boldsymbol{r})$	magnetic field	I.15
h, H	height of surface features	II.3
H_{c1}	magnetic field strength	I.8
$\boldsymbol{H}_d$	demagnetizing field	I.15
$\boldsymbol{H}_{ext}$	external applied field	I.15
H_{ext}	external magnetic field strength	I.8
H_r	remanent magnetic field strength	I.8
I	intensity	I.9.1, I.13
$I(\ldots)$	intensity function	I.3
$I(\boldsymbol{k})$	Fourier transform of image	I.13
$I(x, y)$	image intensity distribution	I.13
I, J, K	crystallographic coordinates	I.13
i, j, k	molecular axes	I.13
I_1	reflected light intensity	I.15
I_{cc}	total current consumption	I.3
I_g	diffracted intensity	I.1
I_R	refractive index	I.7
$\boldsymbol{J}(\boldsymbol{r})$	magnetization vector	I.15
J_c	critical current	I.8

Symbol	Designation	Chapter
J_c	critical end-point dose	I.13
J_e	incident electron dose	I.13
J_1	first-order Bessel function	I.13
J_s	saturation magnetization	I.15
K	elastic constant for curvature	I.13
K	magnetooptical amplitude	I.15
K	proportionality constant	I.1
k	Cliff–Lorimer factor	I.3
$\{\mathbf{k}\}$	disjoint set	I.13
k	integer	I.11
k	inverse decay length for tunneling	I.13
k	Miller index	I.1
$\boldsymbol{k}$	scattering vector	I.13
$\overline{K}$	reciprocal lattice vector	I.1
$\overline{k}$	wavevector of fluctuations	I.5
k_0	$4k/(r+s)$	I.13
k_0	diffracted wave vector	I.1
k_1	statistical factor	I.1
K_{11}, K_{22}, K_{33}	elastic constant of splay, twist, bend	I.13
k_2	factor depending on character of dislocation	I.1
K_A, K_B, K_C	force constant of subchain A, B, C	I.13
k_i	incident wave vector	I.1
K_{Ic}	fracture toughness	I.14
L	distance between parallel walls	I.5
L	length	I.11, I.12, II.3
L	obstacle spacing	I.1
L	sin α	I.9.2
l	characteristic length of a fluctuation wave	I.5
l	Miller index	I.1
l	sample thickness	I.13
L_A	length per unit area	I.12
L_A, L_B, L_C	stretch of subchain A, B, C	I.13
L_L	length per unit length	I.12
L_V	length per unit volume	I.12
M	average distance between boundaries	I.1
M	magnification	I.13
m	mobility	I.13
$\boldsymbol{m}(\boldsymbol{r})$	unit magnetization vector field	I.15
MiTAC	Micro and Trace Analysis center	II.4
M_s	transition temperature	I.1
MSA	methane sulfonic acid	II.4
m_x	magnetization component	I.15
N	number	I.1
N	number of dislocations with measured spacing	II.6
N	number of domain orientations	I.5
N	number of entities	I.12
N	number of pentagons	I.11
N	reflected light amplitude	I.15
N	total number of atoms	I.9.1

Symbol	Designation	Chapter
(n, m)	Hamada indices	I.11
n	director	I.13
n	number of fringes	I.1
n	number of grey scales in an image	I.9.1
n	refractive index	I.7
$N(\boldsymbol{k})$	noise term	I.13
N_A	number per unit area	I.12
N_A, N_B, N_C	number of monomer units in subchain A, B, C	I.13
N_i	number density of molecular dipoles	I.13
N_L	number per unit length	I.12
$O(\boldsymbol{k})$	Fourier transform of object	I.13
P	number of points	I.12
P	polarization	I.13
p	pitch	I.11
p	pressure	I.1
$\overline{p}$	average pressure	I.1
$P(S)$	cumulative probability distribution of S	I.12
P_A	points per unit area	I.12
$P_{ab}(r)$	pair correlation function between atomic species a and β	I.9.1
P_{eq}	equilibrium pressure	I.1
p_f	inter bubble fracture pressure	I.1
p_j	the probability given by an entropymaximizing function (e.g. Gaussian)	I.9.1
P_L	points per unit length	I.12
p_{lp}	minimum loop punching pressure	I.1
P_P	point fraction	I.12
PRC	particle recognition and characterization	II.4
$\boldsymbol{P}_T$	polarizability	I.13
P_V	points per unit volume	I.12
Q	complex magnetooptical material parameter	I.15
Q	observed/expected nearest-neighbor distance ratio	I.12
Q	order parameter	I.5
$\boldsymbol{Q}$	scattering vector, modulus Q	I.9.1
$\boldsymbol{q}$	scattering vector	I.13
q	transmission function	I.13
$q(x, y)$	maximum entropy atom distribution	I.13
$q, \boldsymbol{q}$	modulation vector unreciprocalspace	I.8
Q_e	relative (information) entropy	I.9.1
Q_i	i-th order parameter	I.5
$\boldsymbol{R}$	displacement vector	I.1, I.8
R	observed/expected variance of nearest-neighbor distance ratio	I.12
R	radius of a spherical object	I.1
$\mathbf{R}$	transformation matrix $i, j, k \rightarrow$ I, J, K	I.13
r	length component of the radius vector	I.1
$\boldsymbol{r}$	position vector in real space	I.9.2
r	radius	I.11, I.13
$R(x)$	kinematic intensity ratio	I.3
r_0	dislocation care radius	I.1
r_{eh}	equivalent hole radius	I.1
$\boldsymbol{R}_f$	fault vector	I.1

Symbol	Designation	Chapter
$\boldsymbol{R}_i$	position vector of the ith atom	I.9.1
r_j	position vector of atoms	I.1
R_l	dislocation loop radius	I.1
R_s	particle radius	I.1
S	cos α	I.9.2
S	fiber tensile strength	I.12
s	excitation error (TEM)	I.14
s	shear	I.11
s	tip–sample distance (Section 13.3.2); constant ± ½, ± 1, ± 3/2	I.13
$S(Q)$	structure factor	I.9.1
$s(q)$	structural factor	I.13
s_f	standard deviation of f	I.12
s_g	deviation from the exact Bragg condition	I.1
Sm(α, θ)	speed of step movement	II.3
SR (α, θ)	relative speed of step movement	II.3
T	temperature	I.1, I.5, I.14
T	temperature transfer function	I.13
t	distance in real space	I.13
t	foil thickness	I.1
t	magnitude of the twinning vector	I.8
t	specimen thickness	I.3
t	time	I.5, I.13, I.14
$T(Q)$	contrast transfer function	I.9.1
$T(u)$	phase contrast transfer function	I.1
T_a	annealing temperature	I.5
t_a	annealing time	II.2
T_c	clearing point temperature	I.3
T_c	critical temperature	I.1, I.5, I.8
t_1, t_2	time	II.3
T_m	melting temperature	I.1, I.14
T_t	transition temperature	I.8
t_T	transit time	I.13
U	unreinforced area fraction	I.12
U	voltage	I.14
$\boldsymbol{u}$	difference between diffracted and incident beam	I.1
u	reciprocal coordinate	I.1
u_i	atomic displacement of an atom at the position i	I.5
u_m	reciprocal resolution	I.1
U_h^{obs}	unitary structural factor	I.13
V	applied voltage	I.13
V	number of vertices	I.11
V	volume	I.12
V	volume of the unit cell	I.1
$\boldsymbol{v}$	line direction of dislocation	I.9.2
V(α, θ)	erosion speed in the direction of the surface normal of the inclined step	II.3
V(α, θ, ω, β)	sputtering speed	II.3
$V(\Theta)$	erosion speed in the direction of the surface normal of the average surface	II.3
$(V_V)_r$	reinforcement volume fraction	I.12
V_V	volume per unit volume	I.12

Symbol	Designation	Chapter
W	wall with orientation determined by symmetry	I.5
W	width of a dislocation	I.1
w	Bragg position	I.1
w	wall thickness	I.5
W'	wall with orientation not fully determined by symmetry	I.5
w_g	deviation from the Bragg condition times the extinction distance	I.1
x	element content	I.3
X, Y	Cartesian coordinates	I.7
X, Y, Z	reference coordinates	II.3
x, y, z	coordinates	I.9.1, I.11, II.3
X_i	coordinate on the orthonormal axis of a spontaneous strain tensor	I.5
x_j, x_k	orthonormal coordinates used for the description of atomic displacements and spatial variations of strain patters.	I.5
y	distance of thickness fringes from a cleaved edge	I.3
Z	atomic number	I.3, I.8, I.9.1, I.14
Z	mean nuclear charge	I.9.2
z	beam direction	I.9.1
z	depth	I.1
a	angle	I.11
α	angle of deposition, angle of inclination	II.3
a	proportionality constant	I.1
a^*, β^*, γ^*	reciprocal lattice angles of the low-symmetry phase	I.5
α_0^*, β_0^*, γ_0^*	reciprocal lattice angles of the high-symmetry phase	I.5
a_{ij}	molecular polarizability	I.13
α_{max}	step angle	II.3
β	analyzer angle	I.15
β	angle	I.11
β	angle of the inclined step to the sample coordinate axis x	II.3
β	transition moment	I.13
β_{ijk}	first-order polarizability	I.13
χ	transfer function	I.13
$\chi(u)$	phase shift	I.1
χ_{IJK}	macroscopic second-order hyperpolarizability	I.13
Δ	path difference	I.11
δ	resolution	I.12
ΔF	structure factor difference F_o-F_c	I.13
Δf	defocus	I.1, I.13
Δf	focal length	I.3
ΔG	exess Gibbs free energy for a structural phase transition	I.5
δ_{ij}	unit tensor	I.1
δ_{min}	smallest meaningful value of δ	I.12
δ_{xy}	surface tension	I.13
ε	dielectric constant	I.3
ε	multiplicity factor	I.13
ε	offset parameter	I.8
$\dot{\varepsilon}$	strain rate	I.1
Φ	fluence	I.14
ϕ	angle	I.11

Symbol	Designation	Chapter
ϕ	factor	I.13
ϕ_0	amplitude of the incident beam	I.1
ϕ_c	critical break away angle	I.1
ϕ_c	phased structure factor	I.13
ϕ_g	amplitude of a diffracted beam	I.1
$\phi_g(t)$	amplitude of the scattered beam	I.1
ϕ_h	crystallographic phase	I.13
ϕ_O	slope of helix	I.11
Γ	surface free energy	I.1
γ	angle	I.11
γ	angle of incidence of ion beam with respect to the surface plane	II.3
γ	stacking fault energy	I.11
γ_{APB}	anti phase boundary energy	I.1
γ_{ijkl}	second-order polarizability	I.13
η_i	slip plain normal	I.1
θ	chiral angle	I.11
θ	polar coordinate	I.11
θ	scattering angle	I.9.1
φ	magnetooptical solation angle	I.15
λ	wavelength	I.1, I.9.1, I.11, I.13, II.2
λ'	centre-to-centre spacing	I.1
λ	coupling parameter for two interacting order parameters	I.5
λ_m	characteristic length of patches in tweed and tartan patterns	I.5
Λ_p	effective separation between precipitates	I.1
μ	molecular dipole moment	I.13
μ	shear modulus	I.11
μ_0	vacuum permeability	I.15
ν	Poisson's ratio	I.1, I.11
θ	angle of incidence	I.1, I.15, II.3
ρ	dislocation density	I.1
ρ	polar coordinate	I.11
Σ	the ratio between the volume of a coincidence site lattice and that of the primitive cell	I.1
σ	macroscopic stress	I.5
σ_{ij}	shear component of the stress field	I.1
σ_f	fracture stress	I.1
σ_H	hydrostatic component of the stress field	I.1
σ_h	standard deviation of $U_h{}^{obs}{}_h$	I.13
τ	flow stress	I.1
τ	pulse length	II.2
τ	tan α	I.9.2
τ_0	flow stress from other effects	I.1
τ_{disl}	dislocation component of the flow stress	I.1
τ_{prec}	precipitate component of the flow stress	I.1
τ_{subst}	substitutional component of the flow stress	I.1
ω	'angle the misfit' of domain corners	I.5
ω	angle of rotation between the reference and sample coordinates	II.3
Ω_0	angle	I.11

Symbol	Designation	Chapter
ξ	interaction strength; extinction distance	I.1
ξ_g^{eff}	effective extinction distance	I.1
ξ_g	extinction distance	I.1
x_i	line of intersection of in crystal *i*	I.1
ψ	combination of crystallographic phases	I.13

Abbreviation	Explanation
1-2-3 YBCO	$YBa_2\,Cu_3\,O_7$
1-2-4 YBCO	$YBa_2\,Cu_4\,O_8$
1D, 2D, 3D, 6D	one-, two-, three-, six-dimensional
2-4-7 YBCO	$Y_2\,Ba_4\,Cu_7\,O_{15}$
Al–12Si	Al–12 wt % Si
ALCHEMI	atom location by channeling-enhanced microanalysis
APB	antiphase boundary
appm	atomic parts per million
ASTM	American Society for Testing and Materials
b.c.c.	body-centered cubic
BF	bright field
BMAS	$BaO\text{-}MgO\text{-}Al_2O_3\text{-}SiO_2$
BPSG	borophosphosilicate glass
BSCCO	$Bi_2\,Sr_2\,Ca_n\,Cu_{n+i}\,O_z$
BSE	back-scattered electron
BST	barium strontium titanate
BWR	boiling-water reactor
CA	coaxial
CAI	Ca-Al rich refractory inclusions
CANDU	Canadian deuterium uranium
CD	critical dimension
CIBJO	Confédération Internationale de la Bijouterie, Joallerie et Orfèvrerie
CLSM	confocal laser scanning microscopy
CM	carbonaceous, like the Mighei (meteorite) sub-type
CMC	ceramic matrix composite
CMC	constant mean curvature
CMOS	complementary metal oxide semiconductor
CO	carbonaceous, like the Ornans (meteorite) sub-type
COP	crystal originated particle
CPD	critical point drying
CRT	cathode ray tube
CSF	complex stacking fault
CSL	coincidence site lattice
CTF	contrast transfer function
CVD	chemical vapour deposition
CZ	Czochralski
DBQW	double-barrier quantum well
DBTT	ductile to brittle transition temperature
DEMO	Demonstration fusion reactor

Abbreviation	Explanation
DF	dark field
DHC	delayed hydrogen cracking
DIC	differential interface constant
dpa	displacements per atom
DPTS	demountable polished thin sections
DQC	decagonal quasicrystal
DRAM	dynamic random access memory
DSC	displacement shift complete
DSL	diluted Sirtl etch
EAC	environmental assisted cracking
EBSP	electron backscattering patterns
ECD	equivalent circle diameter
ECP	electron channeling patterns
ED	electron diffraction
EDP	electron diffraction pattern
EFTEM	energy filtering transmission electron microscope
EOR	end of range
EPMA	electron probe microanalyzer
EPRI	electric power research institute
EuTTA	europium thenoyltrifluoroacetonate
f.c.c	face-centered cubic
$FATT_{50}$	fracture appearance transition temperatur
FBR	fast breeder reactor
FE	field emission
FEG	fied emission gun
FET	field effect transistor
FIB	focused ion beam
FIMA	fissions of metal atoms (burn-up nuclear fuel)
FPD	flow pattern defect
FS/RH	Convention to define the sign of the Burgers vector of a dislocation loop: right-handed circuit around a dislocation with F as finish point and S as starting point
FZ	float zone
GIA	Gemmological Institute of America
GP	Guinier–Preston
h.c.p.	hexagonal close-packed
HA	hydroxyapatite
HAZ	heat-affected zone
HDF	high-resolution Faraday
HEMT	high electron mobility transistor
HLW	high-level waste
HOLZ	high-order Laue zone
HRD	HOGE RAAD voor DIAMANT (High Diamond Council)
HWR	heavy water reactor
IAEA	International Atomic Energy Agency
IC	integrated circuit
IDC	International Diamond Council
IDP	interplanetary dust particles
IGA	intergranular attack
IGSCC	intragranular stress corrosion cracking

Abbreviation	Explanation
iPP	isotactic polypropylene
IQC	icosahedral quasicrystal
ITER	international thermonuclear experimental reactor
JJ	Josephson junction
KTA	Kerntechnischer Ausschuss
LC	liquid crystal
LCAO	linear combination of atomic orbitals
LDD	lightly doped drain
LEC	liquid encapsulated Czochralski
LLG	log-likelihood gain
LM	light microscopy
LOCOS	local oxidation of silicon
LPCVD	low-pressure chemical vapour deposition
LPS	long-period superstructure
LRO	long-range order
LTS	low-temperature sensitization
LWR	light-water reactor
MESFET	metal-semiconductor field effect transistor
MMC	metal matrix composite
MOS	metal oxide semiconductor
MPA	Materialprüfanstalt
MPMG	melt powder melt growth
MRI	magnetic resonance imaging
NACE	national association of chemical engineers
NDT	nondestructive test
NEA	nuclear energy agency
OA	optic axis
OBDD	ordered bicontinuous double-diamond
OD	optical diffractogram
OD	optical or apparent depth
ONO	SiO_2-Si_3N_4-SiO_2
OPE	orientational proximity effect
PC	phase contrast
PCI	pellet–cladding interaction
PCP	poorly characterized phases
PEEK	poly(ether ether ketone)
PHWR	pressurized heavy water reactor
PL	photoluminescence
PMC	polymer matrix composite
PMMA	polymethyl methacrylate
ppb	parts per billion
PPO	poly(p-phenylane oxide)
PPS	poly(p-phenylane sulfide)
PSYCO	$Pb_2\,Sc_2\,Ca_{0.5}\,Y_{0.5}\,Cu_3\,O_{8+\delta}$
PTS	polished thin sections
PU	polyurethane
PWR	pressurized water reactor
q+t	quenched and tempered
QC	quasicrystal

Abbreviation	Explanation
R	reflected
r.m.s.	root mean square
RBMK	uranium-graphite channel-type reactor
RD	real depth
RED	reflection electron diffraction
RIE	reactive ion etching
RPV	reactor pressure vessel
SAED	selected area electron diffraction
SAM	scanning acoustic microscopy
SBT	strontium bismuth tantalate
ScanDN	Scandinavian Diamond Nomenclature
SCC	stress corrosion cracking
SCF	self-consistent field
SE	secondary electron
SESF	superlattice extrinsic stacking fault
SISF	superlattice intrinsic stacking fault
sPP	syndiotactic polypropylene
SRC	stress relief crack
SRO	short-range order
SS	stainless steel
TGSCC	transgranular stress corrosion cracking
TIG	tungsten inert gas
TL	thermoluminescence
TR–IMM	transmitted–immersion
TS	tunneling spectroscopy
UV	ultra-violet
VDEh	Verein Deutscher Eisenhüttenleute
VGB	technische Vereinigung der Grosskraftwerkbetreiber
VVER	water-cooled water-moderated energy reactor
YBCO	yttrium–barium–copper–oxygen
YSZ	yttria-stabilized zirconia
ZAF	atomic number, absorption and fluorescence corrections
ZMC	zero mean curvature

Techniques	Explanation
AEEM	Auger electron emission microscopy
AEM	analytical electron microscopy
AES	atomic emission spectroscopy
AES	Auger electron spectroscopy
AFM	atomic force microscopy
APFIM	atom probe field ion microscopy
ARPES	angle-resolved photoelectron spectroscopy
ATRS	attenuated total reflectance spectroscopy
BEEM	ballistic electron emission microscopy
BEES	ballistic electron emission spectroscopy
BF CTEM	bright field conventional transmission electron microscopy

Techniques	Explanation
CBED	convergent beam electron diffraction
CBRHEED	continuous beam reflective high-energy electron diffraction
CCSEM	computer-controlled scanning electron microscopy
CITS	current imaging tunneling spectroscopy
CL	cathodoluminescence
CLSM	confocal laser scanning microscopy
CT	computer-aided tomography
CTEM	conventional transmission electron microscopy
CVD	chemical vapor deposition
DLTS	deep level transient spectroscopy
EBT	electron beam testing
EDS	electron diffraction spectrometry
EDS	energy-dispersive spectroscopy
EDX	energy dispersive X-ray spectroscopy
EELS	electron energy loss spectroscopy
EFTEM	energy filtered transmission electron microscopy
EM	electron microscopy
EPMA	electron probe microanalysis
EPXMA	electron probe X-ray microanalysis
ESCA	electron spectroscopy for chemical analysis
ESEM	environmental scanning electron microscopy
ESI	electron spectroscopic imaging
ESI	element-specific imaging
FFM	friction force microscopy
FIB	focused ion beam milling
FIM	field ion microscopy
FMT	fluorescent microthermography
FT-IR	Fourier transform infrared spectroscopy
HREM	high resolution electron microscopy
HRSEM	high resolution scanning electron microscopy
HRTEM	high resolution transmission electron microscopy
HVEM	high voltage electron microscopy
LACBED	large angle convergent beam electron diffraction
LCT	liquid crystal thermography
LEEM	low-energy electron microscopy
LFM	lateral force microscopy
LM	light microscopy
LMMS	laser microprobe mass spectrometry
LOM	light optical microscopy
LPCVD	low-pressure chemical vapor deposition
LTSLEM	low-temperature scanning laser electron microscopy
M-PIXE	micro-(proton-induced X-ray emission spectroscopy)
MBE	molecular beam epitaxy
MEM	mirror electron microscopy
MFM	magnetic force microscopy
MOVPE	metal-organic vapor phase epitaxy
MRI	magnetic resonance imaging
MULSAM	multispectral Auger microscopy
NMR	nuclear magnetic resonance

Techniques	Explanation
OM	optical microscopy
PCA	principal components analysis
PEELS	photoelectron energy loss spectroscopy
PEEM	photoemission electron microscopy
PFA	principal factor analysis
PIXE	proton induced X-ray emission spectroscopy
PL	photoluminescence
PPM	point-projection microscopy
RBS	Rutherford backscattering spectroscopy
RDE	reactive deposition epitaxy
REM	reflection energy microscopy
REM	reflection electron microscopy
RHEED	reflection high-energy electron diffraction
SAM	scanning acoustic microscopy
SAM	scanning Auger microscopy
SAXS	small-angle X-ray scattering
SCM	scanning capacitance microscopy
SDLTS	scanning deep level tansient spectroscopy
SECM	scanning electrochemical microscopy
SEELS	serial electron energy-loss spectroscopy
SEEM	secondary electron emission spectroscopy
SEM	scanning electron microscopy
SEMPA	scanning electron microscopy with polarization analysis
SFM	scanning force microscopy
SIMS	secondary ion mass spectrometry
SLEEM	slow low-energy electron microscopy
SNOM	scanning near-field optical microscopy
SNPM	scanning near-field probe microscopy
SPE	solid phase epitaxy
SPLEED	spin-polarized low-energy electron diffraction
SPLEEM	spin-polarized low energy electron microscopy
SPM	scanning probe microscopy
SPSTM	spin-polarized scanning tunneling microscopy
SQUID	superconducting quantum interference device
SREM	scanning reflection electron microscopy
STEM	scanning transmission electron microscopy
STM	scanning tunneling microscopy
STOM	scanning tunneling optical microscopy
STS	scanning tunneling spectroscopy
STXM	scanning transmission X-ray microscopy
TED	transmission electron diffraction
TEEM	thermionic electron emission microscopy
TEM	transmission electron microscopy
TL	thermoluminescence
TS	tunneling spectroscopy
TSMFM	tunneling stabilized magnetic force microscopy
TXM	transmission X-ray microscopy
UFM	ultrasonic force microscopy
UMT	ultra microtomography

Techniques	Explanation
VPE	vapor phase epitaxy
WDS	wavelength dispersive spectroscopy
XES	X-ray emission spectroscopy
XPS	X-ray photoelectron spectroscopy
XPS	X-ray photoemission spectroscopy
XTEM	cross-sectional transmission electron microscopy

Handbook of Microscopy
S. Amelinckx, D. van Dyck, J. van Landuyt, G. van Tendeloo

Index

S. Amelinckx, D. van Dyck, J. van Landuyt,
G. van Tendeloo

Handbook of Microscopy

Methods I

Handbook of Microscopy
Applications in Materials Science, Solid-State Physics and Chemistry

Methods I
1997. ISBN 3-527-29280-2.

Methods II
1997. ISBN 3-527-29473-2.

Applications
1997. ISBN 3-527-29293-4.

Further Reading from VCH

S. N. Magonor, M.-U. Whangbo
Surface Analysis with STM and AFM
Experimental and Theoretical Aspects of Image Analysis
ISBN 3-527-29313-2

D. A. Bonnell
Scanning Tunnelling Microscopy and Spectroscopy
Theory, Techniques and Applications
ISBN 3-527-27920-2

Distribution:
VCH, P.O. Box 10 11 61, D-69451 Weinheim (Federal Republic of Germany)
Switzerland: VCH, P.O. Box, CH-4020 Basel (Switzerland)
United Kingdom and Ireland: VCH (UK) Ltd., 8 Wellington Court, Cambridge CB1 1HZ (England)
USA and Canada: VCH, 333 7th Avenue, New York, NY 10001 (USA)
Japan: VCH, Eikow Building, 10-9 Hongo 1-chome, Bunkyo-ku, Tokyo 113 (Japan)

ISBN 3-527-29280-2

Handbook of Microscopy

Applications in Materials Science, Solid-State Physics and Chemistry

Edited by
S. Amelinckx, D. van Dyck,
J. van Landuyt, G. van Tendeloo

Methods I

Weinheim · New York · Basel · Cambridge · Tokyo

Prof. S. Amelinckx
Electron Microscopy for
Materials Science (EMAT)
University of Antwerp - RUCA
Groenenborgerlaan 171
2020 Antwerp
Belgium

Prof. D. van Dyck
Electron Microscopy for
Materials Science (EMAT)
University of Antwerp - RUCA
Groenenborgerlaan 171
2020 Antwerp
Belgium

Prof. J. van Landuyt
Electron Microscopy for
Materials Science (EMAT)
University of Antwerp - RUCA
Groenenborgerlaan 171
2020 Antwerp
Belgium

Prof. G. van Tendeloo
Electron Microscopy for
Materials Science (EMAT)
University of Antwerp - RUCA
Groenenborgerlaan 171
2020 Antwerp
Belgium

Published by
VCH Verlagsgesellschaft mbH, Weinheim (Federal Republic of Germany)

Editorial Directors: Dr. Peter Gregory, Dr. Ute Anton, Dr. Jörn Ritterbusch
Production Manager: Dipl.-Wirt.-Ing. (FH) Hans-Jochen Schmitt

Library of Congress Card No. applied for.

A catalogue record for this book is available from the British Library.

Die Deutsche Bibliothek – CIP-Einheitsaufnahme
Handbook of microscopy : applications in materials science,
solid state physics and chemistry / ed. by S. Amelinckx ... -
Weinheim ; New York ; Basel ; Cambridge ; Tokyo : VCH.
NE: Amelinckx, Severin [Hrsg.]
Methods 1 (1997)
ISBN 3-527-29280-2

Printed on acid-free and chlorine-free paper.

Composition: Alden Bookset, England
Printing: betz-druck, D-64291 Darmstadt
Bookbinding: W. Osswald, D-67433 Neustadt

Short biography of the editors

Severin Amelinckx was born in Willebroek, Belgium, in 1922. He studied at the University of Ghent, receiving his first degree (licence) in mathematics in 1944, his doctorate in physics in 1952, and his aggregation in physics in 1955. Currently he is Emeritus Professor of General Physics and Crystallography associated with the EMAT laboratory of the University of Antwerp (RUCA). Until 1987 he was Director General of the Belgian Nuclear Research Establishment at Mol. He is a member of the European Academy and of the Koninklijke Academie voor Wetenschappen, Letteren en Schone Kunsten van België and former chairman of the division of sciences of this academy.

His research interests include electron diffraction contrast imaging, defects in solids, phase transformations and their resulting domain structures, crystal growth, dislocations, fullerenes and nanotubes, the structure of high-T_c superconductors, modulated structures, and order–disorder in alloys.

Joseph Van Landuyt, who was born in St. Amandsberg, Belgium, in 1938, obtained both his licence (1960) and doctorate in physics (1965) from the University of Ghent. At present he is Professor of General Physics and Crystallography at the University of Antwerp (RUCA and UIA) and of Electron Microscopy at UIA and the Flemish University of Brussels (VUB). He is a member of the Koninklijke Academic voor Wetenschappen, Letteren en Schone Kunsten van België.

His research interests are centered on the study of nanostructural features in alloys, ceramics, and minerals (in particular gems), with special interest in defects in semiconductors and their relation to device performance. More general subjects of interest are structural variants, defects, and phase transitions in various solids.

Gustaaf Van Tendeloo, born in Lier, Belgium, in 1950, received his licence in physics from the University of Brussels (VUB) in 1972, his doctorate from the University of Antwerp (UIA) in 1974, and his aggregation from the University of Brussels (VUB) in 1981. He has been associated with the University of Antwerp (RUCA) since 1972, but has spent extended periods of time as a researcher in Berkeley (USA), Caen (France), and elsewhere. He is currently Professor of Solid-State Physics at the University of Brussels (VUB) and of the Physics of Materials at the University of Antwerp (RUCA and UIA).

His research interests include the electron microscopy of inorganic solids (in particular high-T_c superconductors), solid-state phase transitions, modulated structures, fullerenes, defects in crystals order–disorder in alloys, and nanostructural features in general.

Dirk Van Dyck was born in Wilrijk, Belgium, in 1948. He studied physics, receiving his licence from the University of Brussels (VUB) in 1971 before moving to the University of Antwerp (UIA) for his doctorate (1977) and aggregation (1987). He has been associated with the University of Antwerp since 1971, and is at present Professor of Theoretical Mechanics, Digital Techniques and Image Processing at the University of Antwerp.

Among his research interests are theoretical aspects of dynamic electron diffraction and imaging, holographic reconstruction and structural retrieval, image processing and pattern recognition, and artificial intelligence. In particular, he is involved in the development of a 1 Å resolution microscope in the framework of the Brite/Euram program of the European Union.

The four editors belong to the Electron Microscopy for Materials Science (EMAT) laboratory, University of Antwerp (RUCA), which was founded in 1965. All four have contributed significantly to the development of electron microscopy and its application by numerous publications in journals and books and are members of editorial boards of several international journals in the field of materials science. They have also delivered numerous invited lectures at international conferences in their respective areas of research.

List of Contributors

Amelinckx, Severin (**IV**.1.1.1)
Electron Microscopy for Materials Science(EMAT)
University of Antwerp – RUCA
Groenenborgerlaan 171
2020 Antwerp
Belgium

Baruchel, José (**II**:5)
European Synchrotron Radiation Facility
BP 220
38043 Grenoble
France

Bauer, Ernst (**IV**:1.6)
Physikalisches Institut
Technische Universität Clausthal
38678 Clausthal
Germany

Briggs, Andrew (**III**)
University of Oxford
Department of Materials
Parks Road
Oxford OX1 3PH
U. K.

Cazaux, Jaques (**II**:3)
Université de Reims
Champagne-Ardenne
Laboratoire d'Analyse des Solides Surfaces et Interfaces
(L.A.S.S.I./G.R.S.M.)
U.F.R. Sciences
BP 347 – 51062 REIMS Cédex
France

Cherns, David; Steeds, John W.; Vincent, Roger (**IV**:1.5)
H. H. Wills Physics Laboratory
University of Bristol
Tyndall Avenue
Bristol BS8 1TL
U. K.

Colliex, Christian (**IV**:1.3)
Lab. de Physique des Solides
Université Paris-Sud
Bâtiment 510
91405 Orsay Cedex
France

Courjon, Daniel; Spajer, Michel (**I**:5)
Université de France-Comté
UFR des Sciences et des Techniques
Laboratoire d'Optique P. M. Duffieux
Associe au CNRS, URA 214
Rte. de Gray
25030 Besancon Cedex
France

Cowley, John M. (**IV**:1.2)
Arizona State University
Dept. of Physics & Astronomy
Box 87 15 04
Tempe, AZ 85287-1504
USA

Dhamelincourt, Paul; Barbillat, Jaques (**I**:3)
LASIR, UPR 2631L
CNRS
Université de Lille I
59655 Villeneuve D'Ascq Cedex
France

Fujita, Hiroshi (**IV**:1.4)
Research Center for Science and Technology
Kinki University
Higashi-Osake
Osaka 577
Japan

Huvenne, Jean Pierre;
Sombret, Bernard (**I**:6)
Lab. de Spectrochimie Infrarouge et Raman
Université des Sciences et Technologies de Lilles
Bâtiment C5
59655 Villeneuve d'Ascq
France

Jakubovics, John P. (**IV**:1.7)
University of Oxford
Department of Materials
Parks Road
Oxford OX1 3PH
U. K.

Joyeux, Denis (**II**:4)
Institut d'Optiques Théorique et Appliquée
Centre Universitaire Bât 503
BP 147
91403 Orsay Cedex
France

Lichte, Hannes (**IV**:1.8)
Institut für Angewandte Physik
Technische Universität Dresden
01062 Dresden
Germany

Mouze, Dominique (**II**:2)
Université de Reims
Champagne-Ardenne
Laboratoire d'Analyse des Solides Surfaces et Interfaces
U.F.R. Sciences
BP 347 – 51062 REIMS Cédex
France

Mücklich, Frank (**I**:1) und (**I**:2)
Universität des Saarlandes
Funktionswerkstoffe
Postfach 15 11 50
66041 Saarbrücken
Germany

Petzow, Günter (**I**:1) und (**I**:2)
Max-Planck-Institut für Metallforschung
Heisenbergstr. 5
70569 Stuttgart
Germany

Schlenker, Michel; Baruchel, José (**II**:5)
Laboratoire de Magnétisme
Louis Néel, CNRS, UJF
25, avenue des Martyrs, BP 166
38042 Grenoble Cedex 9
France

Schmahl, Günter (**II**:1)
Forschungseinrichtung Röntgenphysik
Georg-August-Universität Göttingen
Geiststraße 11
37073 Göttingen

Stelzer, Ernst H. K. (**I**:4)
Light Microscopy Group
Cell Biophysics Programme
Meyerhofstr. 1, Postfach 10 22 09
69012 Heidelberg

Van Dyck, Dirk (**IV**:1.1.2)
Electron Microscopy for Materials Science (EMAT)
University of Antwerp – RUCA
Groenenborgerlaan 171
2020 Antwerp
Belgium

Outline

Volume 1: Methods I

Volume 3: Applications

I Classes of Materials

Contents

Volume 1: Methods I

II X-Ray Microscopy

III Acoustic Microscopy

IV Electron Microscopy

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

General Introduction

The importance of microscopic imaging has in recent years been recognized repeatedly by the awarding of Nobel prizes to the inventors of a number of such methods.

As a consequence of the decreasing scale of many devices, high resolution characterization methods have become of vital importance for further development in these areas. Recent advances in data processing have made it possible to develop imaging modes for a number of methods of chemical analysis, based on particle beams; they have been considered as forms of microscopy, particularly as they are often accessories to microscopic equipment.

The systematic development of new materials strongly relies on their characterization at various and increasing levels of resolution. Structure, microstructure, and defect geometry, as well as chemical composition and spatial distribution are important parameters determining the behavior of materials in practical applications.

At present the materials scientist has a large number of methods at his or her disposal to determine these parameters. In applying these methods, use is made of some kind of probe and the response of the sample to this probe is detected and recorded. In many cases the probe consists of a beam of particles such as neutrons, ions, or electrons, or of electromagnetic radiation such as light, X-rays, microwaves, infrared radiation, or sound waves. However the probe may also be a very fine point or fiber in close proximity to the sample surface leading to some form of interaction (mechanical, optical, electrical, magnetic). The probe may be operated in either a stationary or a scanning mode.

As a guiding principle in selecting the characterization methods to be addressed in the Handbook of Microscopy, we used the requirement that the method should give spatially localized information of the microstructure and/or the composition. Moreover, in order to qualify as 'microscopy', the method should have the potential to provide a magnified real-space image of the sample.

An introductory discussion of the physicochemical principles underlying tthe different methods and the type of information which they can provide is the subject of the first two Volumes of the Handbook, *Methods I* and *Methods II*. The chapters have been written by experienced scientists working in the various fields, the main objective being to provide the reader with sufficient insight and information to allow an optimal choice of the method(s) to be used in order to obtain the desired information.

No materials scientist can master more than a few of these techniques, and presumably would not have access to the

instrumentation necessary to apply a number of them. However, on the basis of these descriptions, it should be possible to judge what results can be expected from the various methods and what their limitations are.

Also problems of image recording and data processing are treated in *Methods II*, in relation to the quantitiative evaluation of microscope images in general. The prospective evolution of recently developed methods and the possibilities of new methods still under development are also briefly discussed.

Different classes of materials and different applications of the same material may require different characterization methods; a single method is usually not applicable to all materials. It is therefore meaningful to illustrate the use of the different methods by a number of case studies classified according to the type of material or to its use. This is the objective of the third volume, *Applications*.

The three volumes are complementary. The *Methods I* and *II* volumes mainly address the operator of the instruments and the scientist who wants to understand and interpret the images, while the *Application* volume will be particularly helpful for materials scientists who has to decide which methods looks most promising for their purposes. Together the three volumes form a state of the art account on the subject, essential for the optimal use and application of microscopy techniques.

As the number of available methods grows, increasingly detailed information can be obtained. However, it also becomes increasingly difficult for a materials scientist to make an adequate choice. This Handbook aims to provide a tool to aid in making such a choice possible and to ease access to the vast literature.

The Editors wish to thank the authors for their efforts to provide, in a limited number of pages, clear overviews that are understandable by the nonspecialist.

We are also indebted to the editorial and production staff of VCH for their efficient help in acquiring the manuscripts of so many authors and for turning them finally into finished, well-presented books.

Special thanks are due to Dr P. Gregory, Dr U. Anton and Dr J. Ritterbusch from the Editorial Office for ensuring excellent communications with VCH.

We are grateful for the confidence of the management of VCH in us and for their continuous support.

The Editors:
S. Amelinckx
D. Van Dyck
J. Van Landuyt
G. Van Tendeloo

Part I

Light Microscopy

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

Introduction

Over the roughly 400 years of its development, the light microscope has become an essential device in science and technology. In the field of technical sciences the light microscopes acquired a fundamental status, such as for the investigation of microstructures in materials science, for many tasks in quality control, as well as for the opening of deposits and resources in the fields of mineralogy and geology.

The first investigation of the refraction of light using lenses is documented in the writing of the Arabian *Alhazen*, who experimented with small glassy 'marbles'. The physician P. Borel named the Dutch brothers H. and Z. Jansen as the inventors of the first microscope in the 17th century. Other sources cite C. J. Drebbel from Alkmaar (1572–1634) as the first person to design a light microscope. However, Leeuwenhoek (1632–1723) was very successful in grinding lenses and was able to obtain magnifications of up to 270 times with lenses of a focal length of 1 mm. Using his simple microscope he discovered, for example, the fiber structure of human eye lenses, red blood corpuscles, and bacteria. The Englishman R. Hooke successfully designed an assembled microscope and investigated the structure of plants and cells. He published his results in the *Micrographia* in 1665 [1]. At that time (1669), I. Newton established the emanation theory, assuming light to be a corpuscular radiation. In 1677 C. Huygens contributed the undulation theory, assuming a wave character of light.

Divine improved microscopical imaging by using fixed assemblies of various lenses. Bonanni was the first to attach a distinct illuminator device to the transmission light microscope. Essential steps for further improvement of the instrumentation required the reduction of color aberrations. In contradiction to I. Newton, the possibility of eliminating chromatic defects by the combination of media with various refractions was shown in 1695 by D. Gregory. In 1771 the well-known mathematician L. Euler provided a strong theoretical justification of achromatic optics and initiated the calculation of an achromatic objective. J. von Fraunhofer (1787–1826) started a new era in technical optics with extensive investigations of the properties of optical glass and variations in the composition of glass. Thus, the basis for calculating objectives and oculars became available.

In the 19th century an interactive stimulation drove microscopic development forward together with basic discoveries, mainly in cytology, histology, and bacteriology [2]. One hundred years ago, E. Abbe worked successfully on the theory of image formation in the microscope and derived

the theoretical potential and limit of microscopic imaging [3]. Since the performance of optical systems could be precalculated, the serial development of optimized microscopical systems was taken up, inspired by the new variations of optical glass by O. Schott (1851–1935). The fields of application were further extended by the development of an efficient illumination system (A. Köhler, 1893) and the establishment of new contrasting techniques. H. Siedentopf received the Nobel Prize for the idea of ultramicroscopy, which enabled the detection of particles far below the optical resolution limit. A further Nobel Prize was awarded, in 1953, to F. Zernike for the phase contrast method.

Technological science has profited substantially from the technical availability of polarized light by polarization prisms (Nicol, 1829). Thus, the quantitative and qualitative investigation of crystalline substances were strongly supported. In 1849, H. C. Sorby characterized the birefringence of thin sections of rocks and minerals. Consequently, the investigation of artificial materials also became one of the important fields of light microscopy. H. C. Sorby indicated the importance of the microscopical study of metals. From 1878, A. Martens conceived the metallographic characterization of metals as one of the preconditions for the science of metallography, established by G. Tamann around 1905. Le Chatelier developed the inverted microscope as a distinct technical variation of the microscope for metallographical investigations.

Further developments introduced the exploitation of various other signals and therefore new principles into the field of light microscopy, such as interference microscopy or fluorescence microscopy. Recently, the classical limits of light microscopy have been surpassed by scanning techniques such as laser scanning microscopy or near-field light microscopy.

Although the optical microscope is considered to be a mature instrument, it can be operated with total success only if the basic characteristics of light are understood. Furthermore, the interactions of light with the samples should be clear, otherwise attempts at imaging remain activities of trial and error without optimal contrast results or reliable interpretation. Last but not least, an adequate sample preparation is an indispensable and sometimes underestimated precondition for effective application.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

1 Fundamentals of Light Microscopy

In this Chapter the characteristics of the phenomenon light are indicated (see Sec. 1.1 of this Chapter), which are the basis for various kinds of interaction of light with solid matter (see Sec. 1.2 of this Chapter). The basic knowledge of the characteristics of light, its interaction with the sample, and the origins of imaging (see Sec. 1.3 of this Chapter) are preconditions for the defined exploitation of such interactions, and enable reproducible contrasting of microstructures.

1.1 Light Characteristics

Light is a form of radiant energy absorbed or emitted by spontaneous energy changes of bonding electrons initiating transitions between energy levels in the outer electron shell of an atom. In the electromagnetic theory by Maxwell, light is regarded as superimposed oscillating electric and magnetic fields carrying energy through space in the form of continuous waves. Its behavior is adequately described by Maxwell's equations. According to quantum theory, energy is transported discontinuously in individual bundles called photons. The effects of interaction of light with matter observed in optical microscopy are mainly wave-like in character and can thus be explained by means of wave mechanics.

1.1.1 Wave Propagation

The propagation of waves can be described by a geometrical method, developed by C. Huygens in 1678. Today it is called the *Huygens principle*: every point of an existing wave front is the origin of a new spherical wave which has a propagation velocity and frequency identical to that of the initial wave; the envelope of all such elementary waves leads to the formation of a new wave front (Fig. 1). This is mathematically equivalent to Fermat's principle, which states that light propagates along a path requiring a minimum amount of time.

Electromagnetic waves are transverse waves because the electric vector $\boldsymbol{E}$, also referred to as the electric field strength, oscillates perpendicular to the magnetic vector $\boldsymbol{H}$ (magnetic field strength), and also perpendicular to the direction of propagation x (Fig. 2). For 'normal' light, for example, light emitted by the sun, a candle, or a bulb, the plane of oscillation is not fixed, as the azimuths of vibrations are arbitrary. Devices that

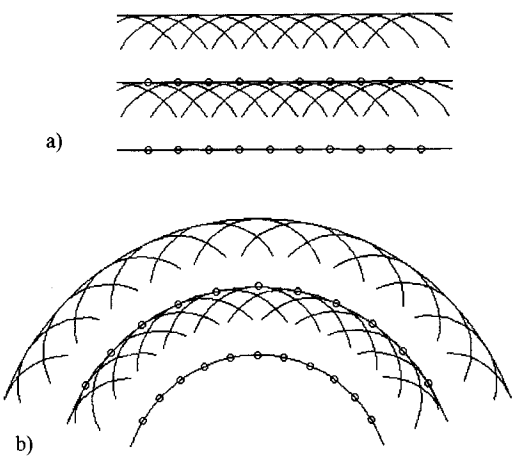


Figure 1. Huygens construction for the propagation of (a) a flat and (b) a spherical wave.

make the vectors vibrate in defined azimuths are called polarizers (see Sec. 1.2.5 of this Chapter).

Light exhibiting one constant plane of oscillation for each vector is linearly polarized or plane polarized (Fig. 2) and possesses a wavelength λ of

$$\lambda = 2\pi \frac{v}{\omega} \tag{1}$$

with v being the phase velocity and ω the angular frequency. The oscillation plane of $\boldsymbol{H}$ is also known as the plane of polarization, and the plane $\boldsymbol{E}$ as the plane of vibration. At any time, both planes are defined by $\boldsymbol{H}_0$ and $\boldsymbol{k}$ or $\boldsymbol{E}_0$ and $\boldsymbol{k}$, respectively, with $\boldsymbol{H}_0$ and $\boldsymbol{E}_0$ being complex vectors (field amplitudes) and $\boldsymbol{k}$ being the wave vector. The source of light can thus be described as a harmonic oscillator causing a time- and space-dependent periodic (sinusoidal) change of the electric and magnetic vector,

$$E = E_0 \, e^{i(kx-\omega t)} = E_0 \, e^{i\omega(x/v-t)}$$
$$H = H_0 \, e^{i(kx-\omega t)} \tag{2}$$

where x represents the actual coordinates and v the phase propagation velocity. The term $(x/v - t)$ then defines the status of oscillation at a given point x at a given time t. This corresponds to a periodic dielectric displacement $\boldsymbol{D}$:

$$D(\omega, t) = \varepsilon_0 E + P = \varepsilon(\omega, t)\varepsilon_0 E(\omega, t) \tag{3}$$

where ε is the dielectric permittivity of a medium, ε_0 the permittivity of a vacuum ($=8.854 \times 10^{-12}\,\mathrm{F\,m^{-1}}$), and $\boldsymbol{P}$ the dielectric polarization. With

$$H = \varepsilon\varepsilon_0 v E \qquad E = \mu\mu_0 v H \tag{4}$$

where μ represents the magnetic permeability and the propagation velocity of an electromagnetic wave (i.e., the rate of phase displacement) is given by the equation

$$v = (\varepsilon\varepsilon_0\mu\mu_0)^{-1/2} \tag{5}$$

Since for a vacuum ε and μ are both equal to 1, the velocity of light in vacuum c can

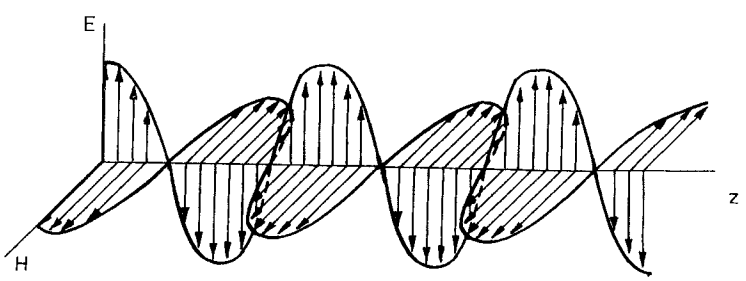


Figure 2. Oscillation of the electrical field (direction E) and the magnetic field (direction H) during the propagation of a polarized electromagnetic wave in the direction z.

be calculated by

$$c = (\varepsilon_0 \mu_0)^{-1/2}$$

$$= 2.997925 \pm 0.0000011 \times 10^8 \,\mathrm{m\,s^{-1}} \quad (6)$$

Conversely, Eq. (5) indicates that the velocity of electromagnetic waves in matter is smaller than in a vacuum since with $\mu \approx 1$ the following equation is obtained as a good approximation for all nonmagnetic media:

$$v = (c/\varepsilon)^{-1/2} \quad (7)$$

This so-called Maxwell relationship between the pure optical measures v, the light velocity c, and the electrical measure ε proves that optical effects can be described by the electromagnetic theory. A generalization of that relationship is, however, not possible, since the phase velocity of light waves depends on their frequency ν, which means that ε is also a function of ν. This behavior is called dispersion, and its explanation can only be a compromise between Maxwell's continuum theory and an atomistic approach. The c/ν ratio is an important characteristic optical constant for matter, and is defined as the refractive index n:

$$n = c/\nu \quad (8)$$

Equations (5) and (8) infer that the refractive index of a material interacting with an electromagnetic wave is a function of the electrical and magnetic properties and chemical composition, as well as a function of the frequency of that particular wave.

1.1.2 Coherence and Interference

Usually light does not consist of a single wavelength (but rather a mixture of wavelengths, that is, 'white' light), and is not plane polarized. The derivation of coherency, interference, and absorption is based on ideally sinusoidal, plane-polarized waves of a distinct wavelength (monochromatic light) (Fig. 2). In a discussion of wave propagation geometry, the wave front concept (Fig. 1) is more likely, and is therefore mostly used to explain reflection and refraction. Since waves vibrate systematically and repetitively, particular points on waves which are in a comparable position—in both space and time—are said to be in the same phase. Sinusoidal waves are in phase if, for example, at an instant in time the crests and troughs are in the same actual or relative positions. A wave front is defined as a surface passing through all points of equal state, that is, of equal phase. Consequently, wavefronts enclose the center of wave initiation (spherical wave fronts). If the source of the waves is point-like and infinitely far away or, alternatively, if it consists of a linear array of single point sources, which is assumed to be the case for the effects observed in microscopy, then the wavefronts are planes. The radius perpendicular to the tangent plane of spherical wavefronts or the line perpendicular to planar wavefronts are called wavefront normals or wave normals. Hence, a wavefront advances in the direction of the wave normal whereas a ray is in the propagation direction of a single wave, that is, the direction of energy flow. Measurement of the light energy yields the intensity I, which is the energy transmitted with time across a unit area perpendicular to the direction of wave propagation. The amplitude A of a simple sinusoidal electromagnetic wave equals the maximum displacement of the electric and magnetic

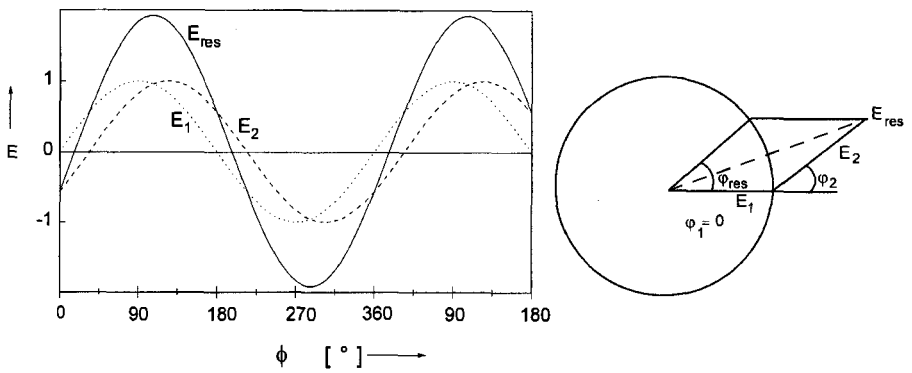


Figure 3. Phase difference between two coherent waves in the $E\phi$ plane. (Adapted from R. Galopin, N. F. M. Henry, *Microscopic Study of Opal Minerals*, W. Heffer and Sons, Cambridge **1972.**)

vectors from their equilibrium positions, and cannot be measured directly because of the high frequencies. However, from the energy of a simple harmonic oscillator it can be shown that $\boldsymbol{I}$ is proportional to A^2 (i.e., $\boldsymbol{E}^2$ and $\boldsymbol{H}^2$).

The phase difference is defined as the difference between the phases of two points on a wave at a given time, or as the difference between the phases of two waves at different times at a given distance from the origin or a fixed reference point (Fig. 3).

The path difference Γ between two waves moving in the same direction along the x axis is the distance of movement in the x direction between comparable points on the waves. The relation of Γ to δ is as follows:

$$\Gamma = \frac{\delta}{2\pi}\lambda \tag{9}$$

where δ is in radians. The interaction between two waves propagating in parallel can be described by the rule of undisturbed superposition, which involves a simple addition of field vectors of the particular waves for all common points in space x_i or in time t_i, respectively. For the components $\boldsymbol{E}_y^{(1)}$ and $\boldsymbol{E}_y^{(2)}$ the following relation holds:

$$\boldsymbol{E}_y = \boldsymbol{E}_y^{(1)} + \boldsymbol{E}_y^{(2)} \tag{10}$$

Consequently, the intensities $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ of the initial waves cannot simply be added—except for a maximum intensity which results at

$$\frac{x_2 - x_1}{\lambda} + \delta_2 - \delta_1 = n \tag{11}$$

with $n = \ldots, -2, -1, 0, 1, 2, \ldots$, because with $\cos 2\pi n = 1$

$$I_{\max} = I_1 + I_2 + 2\sqrt{I_1 I_2} \tag{12}$$

in which particular crests or troughs coincide to give constructive interference,

$$I_{\max} > I_1 + I_2 \tag{13a}$$

For $\cos\pi(m + \frac{1}{2}) = -1$, a crest and a trough coincide with the resulting intensity

$$I_{\min} = I_1 + I_2 - 2\sqrt{I_1 I_2} \tag{13b}$$

(i.e., the interference is destructive: $\boldsymbol{I}_{\min} < \boldsymbol{I}_1 + \boldsymbol{I}_2$).

In a particular case, when $\boldsymbol{I}_1 = \boldsymbol{I}_2$, Eq. (13) yields $\boldsymbol{I}_{\max} = 4\boldsymbol{I}_1$ and $\boldsymbol{I}_{\min} = 0$, respectively. Since according to the rule of energy conservation, intensity cannot simply be

eliminated, every point in space and time with $I_{max} > I_1 + I_2$ corresponds to another point with $I_{min} < I_1 + I_2$. Consequently, interference can be defined as a spatial redistribution of energy.

Of course, local intensity extinctions can be observed only if they remain at a constant position during the observation time. This is equivalent to the condition that $\delta_2 - \delta_1$ is constant with time, that is, that light which oscillates along individual rays is in phase or has a constant phase difference yielding a systematic and periodic interference. Waves behaving in this way are said to be coherent. Only coherent waves can cause interference. For all other conditions which are present in the majority of cases under normal conditions the particular intensities have to be added.

1.2 Light Interaction with Solids

The wavelength of visible light is three orders of magnitude higher than the spacings between the atoms and ranges between approximately 360 nm (violet light) and 780 nm (red light). Therefore, for light microscopy, in comparison to other methods of material characterization, based on electromagnetic waves, the interaction with solids can be dealt with as a continuum effect, where the solid matter consists of three-dimensional arrangements of distinct atoms and thus represents a microscopically discontinuous structure (Fig. 4).

In the following section, some remarks will be made concerning basic questions such as:

- Why are some materials transparent to visible light?
- Why does graphite appear black?
- Why are metals shiny?

The optical effects discussed afterwards are *reflection*, that is, the interaction of light with an optical interface, where the penetration of the media forming the interface is commonly neglected; *refraction*, which deals with the bending of the propagation direction of light on passing an optical interface and penetrating a

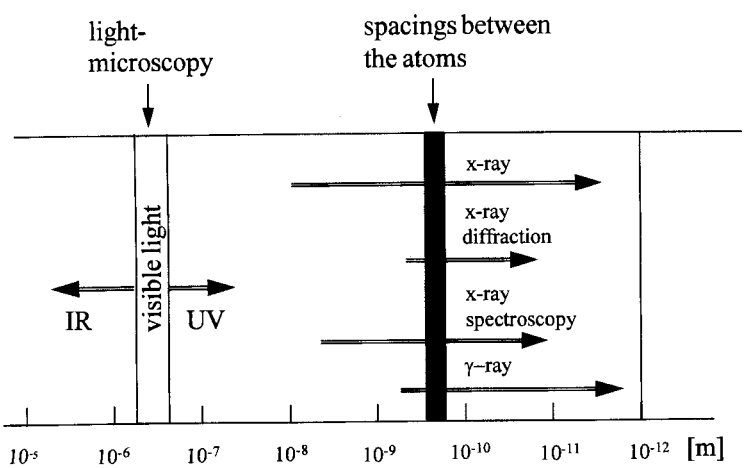


Figure 4. Dimensions of electromagnetic waves, used for the microstructural characterization of solids.

medium; *absorption*, which involves the conversion of light energy into other types of energy; and finally the consequences of reflection, refraction, and absorption as well as scattering on the polarization of electromagnetic waves.

1.2.1 Remarks on Transparency, Luster, and Color of Matter

The answers concerning transparency, luster, and color of matter are all based on the fact that the electrons of atoms are forced to undergo tiny vibrations when they are exposed to light. The amplitudes of these vibrations are not more than 10^{-17} m or less than 1% of the radius of an atomic nucleus. Nevertheless, all light and colors from the objects around us—as well as from the objects seen with the microscope—are produced by these small vibrations under the influence of sunlight or artificial light.

The cloud of electrons of each atom vibrates under the influence of light, which is normally a mixture of many frequencies. The motion of an oscillator exposed to such a mixture of frequencies is simply a superposition of all the motions that it would perform if exposed separately to the light of each separate frequency contained in the mixture.

If an electromagnetic wave of frequency ν passes over an electron oscillator, the electric field exerts a periodic force and leads to certain characteristic responses. First of all the periodic electric field induces a vibration of the oscillator so that it oscillates with the frequency of the field, not with its own resonance frequency ν_0. The amplitude and the phase of this motion depend on the relative values of ν and ν_0. If ν is much smaller than ν_0, the oscillation is weak and in phase with the driving electric force of the light. If ν is much larger than ν_0, it is also weak but opposite in phase to the driving force. If ν is in resonance (in which case ν equals ν_0), the oscillation is stong and out of phase.

Crystalline Matter

In order to understand the effect of light on matter in bulk, it must be explained how electromagnetic waves react to a large number of more or less regularly arranged oscillators, when the average distance between the oscillators is small compared with the wavelength of visible light. As mentioned above, every oscillator emits a light wave. If these oscillators are more regularly arranged, their waves tend to interfere with one another in a definite way (see Sec. 1.1.2 of this Chapter). If light impinges on a surface of a solid object, the oscillators (i.e., the electrons of atoms) vibrate under the influence of the incident light and emit light waves. In the bulk of the material all these light waves, apart from some weak incoherent scattering due to irregularities and imperfections, add up to one strong refracted wave. However, near the surface of the material, there is no adding up because of the existence of a thin layer of oscillators at the surface (known as the first Fresnel zone, about as deep as $\lambda/2$) for which the back radiation is not completely cancelled by interference. The radiations behind these oscillators add up to a reflected wave. The intensity of the reflected light is practically independent of the frequency. Therefore, an incident mixture of white light is reflected as a

'white' composite of many individual radiations. This explains the reflectivity and transparency of, for example, water, quartz glass, and quartz crystals. For the last example, the oscillators, since they are not randomly distributed, reradiate in unison, that is, the reflected light is coherent.

Substances with Transparency

If light impinges on these substances, it is partially reflected at the surface without preference for any color. The rest of the light enters the substance and propagates as a refracted wave within it. Therefore, these objects appear colorless. Their outlines are nonetheless visible because of the reflection of the light at the surfaces. Sometimes such objects may exhibit color under special circumstances—reflection and refraction are only approximately independent of frequency. Both increase slightly at higher frequencies because such frequencies are a little closer to the natural resonance of the atom.

If colorless substances are in the form of small grains, each grain being larger than the wavelength of light, the substances appear white. The incident white light is partially reflected in many directions, depending on the orientation of the grain surfaces. The light that penetrates the grain is again partially reflected on the inside surfaces, and after several reflections and refractions it returns to the eye of the observer from various directions. Since none of these processes discriminates against any color, the returning light will be white and diffuse. If the grains are smaller than the wavelength of light, there are not enough oscillators in the grain to establish ordinary reflection and refraction. The situation is then similar to that in a gas of independent molecules in which the light is scattered [4].

Colors

Most of the colors we see are due to preferential absorption (see Sec. 1.2.4 of this Chapter). Most atoms and molecules have resonances only at infrared and the ultraviolet wavelengths. In order to produce a resonance in the visible region, the excitation energy must be between 0.5 and 3 eV. These values are rather small for electron excitations, but large for molecular vibrations. There are, however, atoms and molecules having excited states in that region. These are atoms with several electrons in incomplete shells and certain organic compounds, the dyestuffs. Such atoms can be excited by rearranging the electrons in the incomplete shell, which requires less energy than excitation to a higher shell. In these terms, objects are black when there is absorption for all visible frequencies. Black objects do not absorb all falling light; some reflection always exists at the surface. Thus, reflection occurs at a polished black surface, such as the surface of a graphite microsection. A dull black surface reflects as strongly as a polished one, but the reflected rays are distributed in all directions. However, the fact that the graphite appears black is also attributed to the fact that in such weak conducting materials the lower concentration of free electrons leads to lower absorption as well as a deeper penetration of incident waves. Therefore, the reflected waves also come from various depth regions and are able to extinguish each other [5].

Table 1. Optical and electrical properties of some materials

Phase	Crystal system	n (for Na_D, $\lambda = 589\,nm$)	R (%)	$1/\rho(\Omega^{-1}m^{-1})$
Ag	Cubic	0.181	95	630.1×10^5
Au	Cubic	0.366	85	451.1×10^5
Cu	Cubic	0.64	80	595.9×10^5
Hg	Liquid	1.73	78	10.6×10^5
Sb	Cubic	3.0	74	25.6×10^5
Pt	Cubic	2.06	70	95.2×10^5
α-Fe	Cubic	2.36	56	48.0×10^5
C	Cubic	2.43	17–45	0.37
C	Hexagonal	1.1	3–25	1.25×10^5
BN	Hexagonal	1.71–1.72	7.5	1×10^{-11}
TiO_2	Tetragonal	2.61	21–25	0.034–0.0011

Metals

A metal is characterized by the fact that within the confines of the material there are many electrons—the conduction electrons—extending over many atomic diameters. These electrons are most important for the optical properties of metals (Table 1).

There are one, two, or sometimes three electrons per atom among the conducting electrons. The rest of the atomic electrons remain bound to the atoms. The conduction electrons can be regarded as an electron gas which penetrates the crystal lattice without much hindrance. Its motion is not completely free; however, it can be considered as freely moving for the reaction of these electrons with visible light. There is one important difference between the impinging of light on a metallic surface and light striking the surface of a liquid or a crystal. Since the resonance frequencies of a liquid or a crystal are higher than the frequency of light, they vibrate in phase with the light. In a metal, however, the electrons are in opposite phases. Under these conditions the refracted light wave cannot be propagated if the density of the electrons and the amplitude of their vibration are above a certain limit. The limit can be expressed in terms of the 'plasma frequency' ν_p, which is given by the equation

$$\nu_p = (N_e e^2/m_e)^{1/2} \tag{14}$$

where N_e is the number of electrons per cubic centimeter and m_e is the electron mass. This frequency is usually in the ultraviolet. Whenever the light frequency is less than ν_p, as is always the case for visible light, no refracted wave can develop in the medium; there are too many electrons inside the medium that are moving in phase opposite to the light. Therefore, all the energy of the incoming light must go into the reflected wave. As discussed above, this reflected light is produced in a thin layer at the surface of the metal, a layer no thicker than the wavelength of the light. A more exact calculation shows that in a metal this thickness is equal to the wavelength corresponding to the plasma frequency divided by 2π. However, the wave reflected from a metal surface has almost the full intensity of the incoming wave, apart from small energy losses due to the already mentioned 'friction' of vibrating electrons in the surface. This is

why 'white' metals such as silver and aluminum are so shiny: they reflect almost all visible light regardless of its frequency. Smooth surfaces of these metals are therefore ideal mirrors. In colored metals such as copper or gold there are additional losses. These losses come from absorption by electrons other than the conduction electrons. In copper and gold, the bound electrons are part of an incomplete shell and have resonances in the blue–violet region that leads to absorption. As a result, copper and gold have a reddish yellow appearance [6].

1.2.2 Reflection

According to the fundamental law of reflection, the angles of incidence and reflection, measured relative to the normal of the reflecting surface, are equal and in the same plane as the plane of incidence. Incident light waves with a common wavefront generate point sources of reflected waves which also possess a common wavefront. The incident wavefront AA′ strikes the surface plane first at position A (Fig. 5). During the propagation of the wave from position A′ to the other surface position B the elementary wave from A propagates to B′. Consequently, the angle of the incident wavefront ϕ_i and the angle of reflection ϕ_{refl} are equal.

In the case of flat, accurately polished surfaces a regular reflection is obtained (surface roughness ≪ wavelength), whereas rough surfaces exhibit a diffuse reflection, as all hills and ridges act as distinct mirror planes (surface roughness ⩾ wavelength). The intensity and wavelength of the reflected beam depends on the nature of the surface and the optical properties of the reflecting medium. Complications arise if a material is more or less opaque to certain wavelengths (selective absorption). Moreover, all graduations exist between transparent and opaque matter of metallic and nonmetallic luster (see Sec. 1.2.1 of this Chapter). Since for the microscopic examination of materials in reflected light a vertically incident beam is usually used, the following discussion is limited to surface orientations normal to the incident beam. The theory and the methods have been described in detail elsewhere [7–10].

The ratio of the intensity of the reflected beam I_{refl} to the incident beam I_i is defined as the reflectivity R, which is

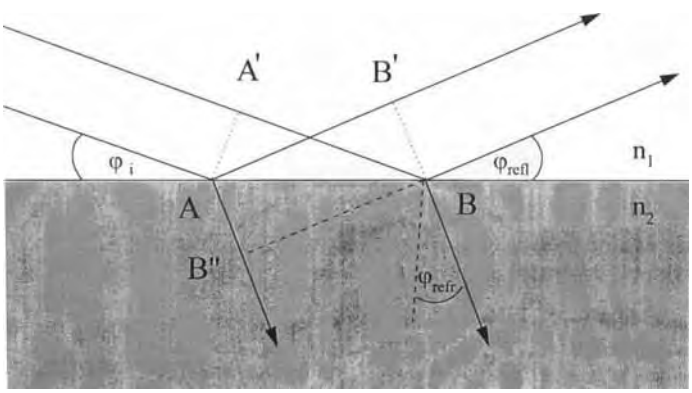


Figure 5. Geometrical representation of Huygens' principle for the explanation of the fundamental laws of reflection and refraction.

given by

$$R = I_{\text{refl}}/I_{\text{i}} = \frac{(n-n_0)^2 + k^2}{(n+n_0)^2 + k^2}$$

$$= \frac{(n-n_0) + n^2\chi^2}{(n+n_0) + n^2\chi^2} \qquad (15)$$

where n is the refractive index, k the absorption coefficient, χ the index of absorption in the tested medium, and n_0 the corresponding refractive index of the medium in which the light propagates before entering the absorbant (see Sec. 1.2.4 of this Chapter). In air, n_0 is equal to 1.00029 (Na line, 1 atm, 20°C), and in a vacuum n_0 is exactly equal to 1, which gives the following relation for both media without major error:

$$R = I_{\text{refl}}/I_0$$

$$= \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2} = \frac{(n-1) + n^2\chi^2}{(n+1) + n^2\chi^2} \qquad (16)$$

For transparent materials k is close to zero, which simplifies Eq. (16) to

$$R = \frac{(n-1)^2}{(n+1)^2} \qquad (17)$$

In a similar way, a coefficient of transmission τ is attributed to the intensity ratio of an incident ray and a refracted ray. For coupled reflection and refraction at the interfaces of transparent media, $R + \tau = 1$.

Like n or k, R is also a function of dispersion of the reflectivity and depends on the electrical conductivity. In anisotropic substances, R varies with the orientation of the polarization plane relative to the symmetry elements (reflection anisotropy, bireflectance). Examples of materials which exhibit strongly anisotropic behavior are graphite (carbon, hexagonal crystal system), molybdenite (molybdenum sulfide (MoS_2), hexagonal crystal system), and covellite (copper sulfide (CuS), hexagonal crystal system). Reflectivity is measured on carefully polished samples by a photometer. Monochromatic light of at least three different wavelengths must be used. For calibration, a standard material of known R and k is measured, such as a single crystal of tungsten carbide or silicon carbide.

n and k can be obtained from Eq. (16) if the reflectivity is measured as a function of two immersion media for a given wavelength. If R' of a sample is determined in a medium of known refractive index n_0', a second examination with another medium of known n_0'' is carried out to give R''. n is then calculated according to

$$n = \frac{n_0''^2 - n_0'^2}{2[n_0''(1+R'')/(1-R'') - n_0'(1+R')/(1-R')]} \qquad (18)$$

For air ($n_0' \approx 1$), Eq. (18) gives

$$n = \frac{n_0''^2 - 1}{2[n_0''(1+R'')/(1-R'') - (1+R')/(1-R')]} \qquad (19)$$

The absorption coefficient k is obtained from

$$k = [(n+n_0'')^2 R'' - (n-n_0'')^2]^{1/2} \times (1-R'')^{-1/2}$$

$$= [(n+n_0')^2 R' - (n-n_0')^2]^{1/2} \times (1-R')^{-1/2} \qquad (20)$$

Some reflectivity data for isotropic and anisotropic materials are listed in Table 1. Silver is the brightest phase when in

polished sections, followed by other noble metals (for an explanation, see Sec. 1.2.1 of this Chapter). Other metals exhibit 60–70% reflectivity, arsenides between 50 and 60%, followed by complex sulfides (40–50%) and antimonides (30–40%). Oxides and some sulfides such as ZnS, MoS_2, and Ag_2S exhibit 10–30% reflectivity. Extensive tables exist for ore minerals, but only a small amount of data is available for intermetallic phases and alloys.

The general relation between n and R according to Eq. (17) exhibits a minimum for $n = 1$ and approaches $R = 1$ (100% reflectivity) for $n = 0$ and $n = \infty$. If absorption and effects due to surface roughness are neglected, a certain luster can be attributed to a particular portion of this curve. Materials with $n = 1.3$–1.9 and $R < 0.09$ (e.g., halogenides, carbonates, sulfates, phosphates, arsenates, silicates, and some oxides) show a glassy luster. A diamond-like luster is attributed to materials with $n = 1.9$–2.6 and R between 0.09 and 0.20 (diamond, SnO_2, TiO_2, ZrO_2, ZnS, sulfur). The luster at n ranging between 2.6 and 3.0 and $R = 0.20$–0.25 is considered semimetallic (e.g., Ag_2S, HgS, Cu_2O, and Fe_2O_3). Materials with $n > 3$ and $R > 0.25$ possess a metallic luster (MoS_2 and other sulfides and arsenides, metal carbides, metal borides, most of the metallic elements, alloys, and intermetallic phases). Gold and silver, however, are examples of elements with a metallic appearance at $n < 1$ (see Table 1). Platinum is an exception from Eq. (17), since for $n = 2.06$ a reflectivity of 12% instead of 70% is expected. This is due to the fact that absorption has to be taken into account as an additive term according to Eq. (16).

1.2.3 Refraction

Light waves are bent or refracted on passing from one transparent medium to another when the densities or optical properties of the media differ. Thus, considering Fig. 5, during the propagation of the wave from position A′ to the surface position B the elementary wave from A also propagates to B″ in the stronger refracting medium ($n_2 > n_1$). The capability of a material to refract light is called refringence. The angle of refraction ϕ_{refr} is defined as the angle between the refracted ray and the normal to the interface between the media (see Fig. 5). ϕ_{refr} is a function of the angle of incidence ϕ_i as well as the wavelength λ, or in other words, a function of the velocity v in the particular dielectric medium. At the interface between isotropic media the geometric situation is determined by Snellius' law:

$$\frac{v_i}{v_{refr}} = \frac{\lambda_i}{\lambda_{refr}} = \frac{\sin \phi_i}{\sin \phi_{refr}} = \frac{n_{refr}}{n_i} \qquad (21)$$

where n_{refr} is the refractive index of the refracting medium, and n_i is the refractive index of the medium in which the beam propagates prior to incidence (e.g., air).

According to Snellius's law there is no refraction at an incidence normal to the interface. If light passes from a transparent medium of high n_i into one of lower n_{refr}, a critical angle of incidence ϕ_{tot}, exists at which the light is totally internally reflected at the interface (total reflection):

$$\text{for} \quad n_i \gg n_{refr} \rightarrow \frac{n_i}{n_{refr}} = \sin \phi_{tot} \qquad (22)$$

The effect of total reflection is of particular importance for the determination of refractive indices, for the construction of

optical instrumentation as well as for the identification of pores and inclusions in transparent materials (see Sec. 1.3.1 of this Chapter).

Since in isotropic media the dielectric properties at optical frequencies are given by Eq. (3), the light velocity v_{refr} is equal in all directions, and thus Eq. (3) holds for all ϕ_i. The numerical operation of v_{refr} into the directions of light transmission, known as the ray velocity tensor surface, then results in a sphere. The bending of the wavefront and ray direction can be derived using Huygens's construction (Fig. 6, where EC_1 is the refracted wavefront, similar to the isotropic case).

In an anisotropic medium, however Eq. (3) has to be replaced by inserting the second-rank dielectric constant tensor ε:

$$D_i = \varepsilon_0 \varepsilon_{ij} E_j \tag{23}$$

This means that the light velocity now depends on the transmission direction and that $\boldsymbol{D}$ has a different direction to $\boldsymbol{E}$ (Fig. 6).

As a result of Maxwell's equations, two waves of different velocity and perpendicular polarization plane propagate

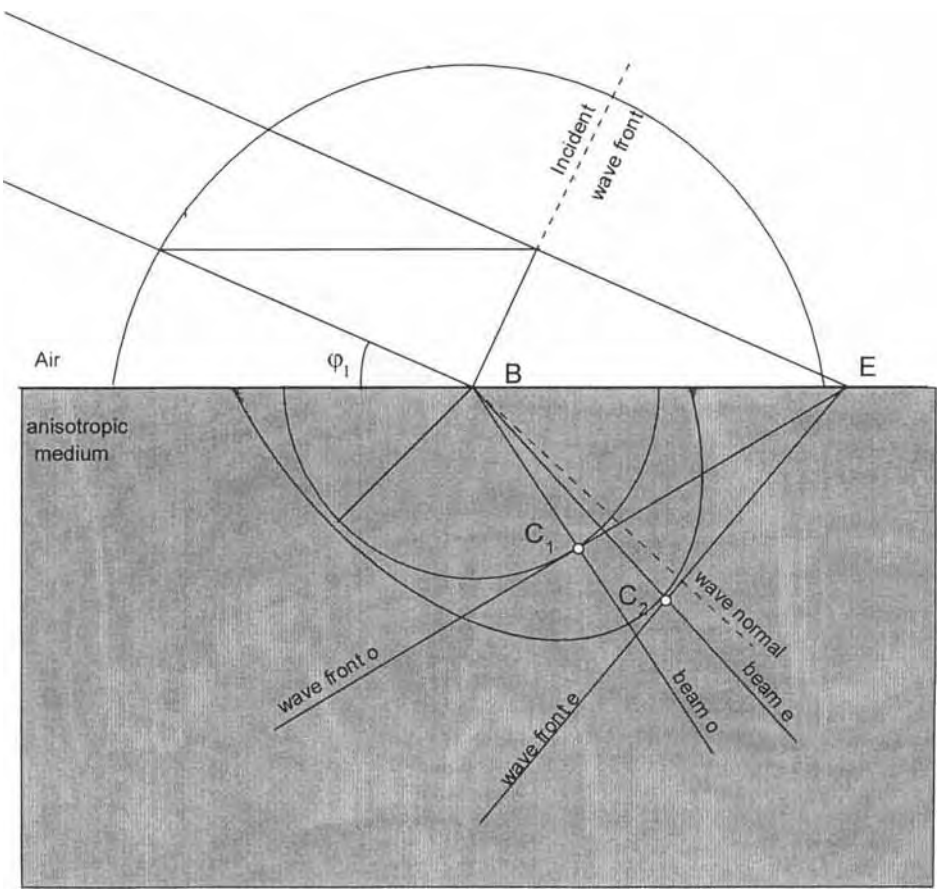


Figure 6. Huygens construction of anisotropic refraction using velocity surfaces.

through the anisotropic crystal. One of these waves behaves as in the isotropic case (Fig. 6, EC_1), and is therefore referred to as an ordinary wave or ordinary beam, o, whereas the other is called extraordinary, e. Beam splitting is known as double refraction or birefringence. The geometric representation of the velocity of the ordinary beam v_e as a function of the transmission direction gives an ellipsoid known as the extraordinary ray velocity surface, which touches the sphere of the ordinary beam velocity v_0 at a circular cross-section. The effect of wave splitting can also be illustrated by Huygens's construction. In Fig. 6, beam e propagating from B to C_2 is faster than beam o propagating from B to C_1, and hence creates a wavefront with a different angle of refraction. Hence, Snellius's law is valid for the ordinary wave but not for the extraordinary beam. In addition to that, which means the wavefront e is not perpendicular to its propagation direction (beam e), the wave normal and the ray direction are no longer parallel. Although the deviation is small, it results in an ovaloid for the wave normal instead of an ellipsoid for the ray direction.

Since the representation of v_0 and v_e by using two interconnected index surfaces is not very comprehensive and does not contain any information about the polarization directions, another kind of figure was introduced by Fletcher (1892, based on the ideas of Fresnel) called the indicatrix, which shows simultaneously the direction dependence of both refractive indices as well as that of the particular polarization mode. The indicatrix is defined as the tensor surface of the reciprocal dielectric tensor or the tensor of dielectric impermeability. Its relationship to the crystal system is given in a system of principal axes x_i $(i = 1, 2, 3)$ by

$$\frac{x_1^2}{\varepsilon_{11}} + \frac{x_2^2}{\varepsilon_{22}} + \frac{x_3^2}{\varepsilon_{33}} = 1$$

and for $n_i^2 = \varepsilon_{ij}$

$$\frac{x_1^2}{n_1^2} + \frac{x_2^2}{n_2^2} + \frac{x_3^2}{n_3^2} = 1$$

or

$$B_1 X_1^2 + B_2 x_2^2 + B_3 x_3^2 = 1 \qquad (24)$$

with ε_{ij} being the principal dielectric constants and the quantities $(B_i = 1/n_i^2)$ the relative optical-frequency dielectric impermeabilities [11].

The indicatrix can thus form an ellipsoid with the principal axes $\boldsymbol{n}_1 < \boldsymbol{n}_2 < \boldsymbol{n}_3$. The symmetry, shape, and orientation of the indicatrix relative to the crystal are defined by the crystal symmetry, the chemical composition, the wavelength, and the temperature. Generally, for a cubic, that is, an isotropic, crystal, the indicatrix becomes a sphere. Since all sections are circles, there is no beam splitting, no direction of preferential polarization, and hence no birefringence. The size of the sphere and thus the principal refractive index n is solely defined by the chemical composition and the character of the atomic bonds. Since the symmetry of the indicatrix must be compatible with that of the crystallographic system in trigonal, tetragonal, and hexagonal systems, the indicatrix is necessarily an ellipsoid of revolution about the principal symmetry axis c (see Fig. 7).

Hence, there are only two principal refractive indices n_1 and n_3. The central section perpendicular to the principal axis c, and only this section, is a regular circle (Fig. 7), where o refers to an 'ordinary

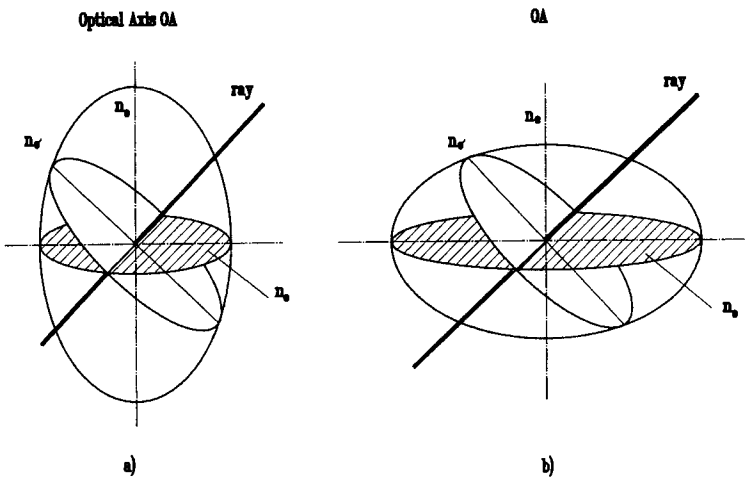


Figure 7. Uniaxial indicatrix: (a) optically positive and (b) optically negative (planes of circular sections are indicated). For an explanation, see the text.

beam'. For a wave propagating along the principal axis c there is no double refraction, which means that this direction behaves isotropically. This preferential axis of isotropy is called the optical axis (OA), and the crystal is said to be uniaxial. The propagation direction of the extraordinary beam with refraction index n_e is thus perpendicular to the optical axis and vibrates in parallel to it. The crystal is said to exhibit an optically positive character '(+)' for $n_e - n_o > 0$ (i.e., if $n_e = n_3$, $n_o = n_1$) and an optically negative character '(−)' for $n_e - n_o < 0$ ($n_o = n_3$, $n_e = n_1$). An optically positive uniaxial indicatrix is thus elongated in the direction of the optical axis, whereas a negative indicatrix appears compressed (Fig. 7).

In the orthorhombic, monoclinic, and triclinic systems the indicatrix is a triaxial ellipsoid. There are two circular sections and hence two privileged wave normal directions of quasi-isotropy in which there is no double refraction. These two directions are called the optical axes, and the crystal is said to be biaxial (Fig. 8).

In positive biaxial crystals, n_β approaches n_α, the indicatrix therefore approaches the shape of a prolate ellipsoid of revolution, that is, the form of the uniaxial positive indicatrix. In the negative biaxial case, as n_β approaches n_γ, the indicatrix assumes the shape of a negative uniaxial indicatrix with a horizontal optical axis OA (Fig. 8). The circular sections always pass through the Y axis, called the optical normal direction, which means that n_β is always the refractive index of the quasi-isotropic direction. The relative positions of the optical axes are hence a function of n_α and n_γ since they are located in the XZ plane, also known as the plane of optical axes or optical plane with n_β normal to it. Consequently, the optical character of the biaxial indicatrix can also be derived from the angle between the optical axes. The smaller (acute) angle is known as the optical angle $2V$. In positive crystals, $2V$ is bisected by the vibration direction of n_γ, that is, the Z axis of the indicatrix, which is then called the acute bisectrix, $2V$. In negative crystals the vibration direction of n_α, that is, the X axis, divides $2V$, which means that the Z axis is now the obtuse bisectrix. The optical angles are usually specified as $2V_x$

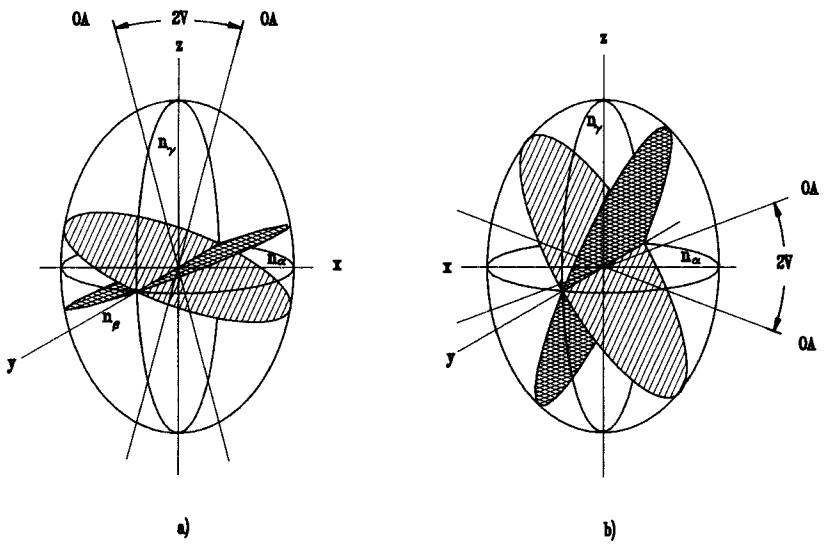


Figure 8. Biaxial indicatrix: (a) optical positive and (b) optical negative (planes of circular sections are indicated).

and $2V_z$, and can be measured directly by means of conoscopy (see Sec. 2.1 of this Chapter) or calculated by

$$\cos^2 V_x = \frac{n_\gamma^2}{n_\beta^2}\left(\frac{n_\beta^2 - n_\alpha^2}{n_\gamma^2 - n_\alpha^2}\right)$$
$$\tan^2 V_z = \frac{n_\gamma^2}{n_\alpha^2}\left(\frac{n_\beta^2 - n_\alpha^2}{n_\gamma^2 - n_\beta^2}\right) \qquad (25)$$

or, for a material of small double refraction, by Mallard's approximation,

$$\tan V_z \approx \left(\frac{n_\beta - n_\alpha}{n_\gamma - n_\beta}\right)^{-1/2} \qquad (26)$$

Generally, $2V$ is a very sensitive measure for small variations in symmetry, chemical composition, or mechanical stresses.

The crystallographic position of the indicatrix relative to the crystal axes depends on the crystal symmetry. In the orthorhombic system the principal axes of the indicatrix are always parallel to the principal crystallographic axes, which also fit into the symmetry elements of both figures. In the monoclinic system, one principal axis of the indicatrix must be parallel to [010] of the crystal, whereas the inclination of the other axes relative to [100] or [001] is free. This means that both the indicatrix and crystal share a common mirror plane. In the triclinic system there is no correlation between the indicatrix and crystal axes at all. In this case the position of the indicatrix is determined only by the type, concentration, and position of the atoms in the crystal structure interacting with the light, which means it is a characteristic material constant.

The indicatrix only exists, of course, as a well-defined description of the optical behavior of a material for a particular wavelength. The dependence of n, $2V$, and the orientation upon λ is called dispersion. In colorless materials the refractive indices decrease with increasing wavelength (normal dispersion). The

slopes of the functions n_α, n_β, and n_γ versus λ may be equal, convergent, or divergent. This means that for uniaxial systems the shape of the indicatrix varies with λ, whereas the direction of the optical axis may not. In the orthorhombic system $2V$ varies in relation to λ, and may even approach uniaxiality for a particular λ with a subsequent change of the optical character (i.e., rotating the optical plane by 90°). In monoclinic crystals the indicatrix may also rotate about the Y axis (optical normal), as it is parallel to the crystallographic [010] direction, which creates a change in the inclination angle between the other principal and crystallographic axes (inclined dispersion). Crossed dispersion occurs if one of the principal axes of the indicatrix is the common acute bisectrix for all wavelengths and if it is parallel to [010] while the other axes are freely oriented. The case of the obtuse bisectrix X of all colors being parallel to [010], and the acute bisectrix Z as well as the optic normals having different orientations for different wavelengths, is known as horizontal or parallel dispersion. In the triclinic system the behavior of the indicatrix in relation to λ is irregular.

For a complete description of the optical properties of a material a set of constants has to be known: n_α, n_β, n_γ, Δ, $2V_z$, and the optical character given as (+) or (−), which determines the shape and size of the indicatrix. The relative orientation of the indicatrix is usually given in terms of the angle between the principal axes of the indicatrix and the crystallographic axes, for example, $X = c$, $Y = b$, $Z = a$ (orthorhombic system) or $X \wedge c = 20°$, $Y = b$ (monoclinic system). The dispersion is described by the relation between $2V_{red}$ and $2V_{violet}$ in terms of $2V_{red} < 2V_{violet}$, $2V_{red} \gg 2V_{violet}$, etc. or in abbreviation form $2V_{red} < \nu$, $2V_{red} \gg \nu$, respectively. If necessary, information on colors may also be added, for example, X = pale green and Z = dark green (see Sec. 1.2.4 of this Chapter).

1.2.4 Absorption

All rules of light propagation can be derived by Maxwell's theory, but the quantum theory gives explanations for the effects of emission and absorption, which is the transformation of wave energy to other types of energy, for example heat energy. Particular wavelengths usually excite optical electrons of the outer shells of an atom and are eliminated from the set of frequencies of the incident beam. As a result, the absorbing materials appear colored, and the intensity of the incident beam is weakened. The absorption is a resonance effect. An oscillating electric field ($\boldsymbol{E} = \boldsymbol{E}_0 \cos \omega t$) applies a force ($e\boldsymbol{E}_0 \cos \omega t$) on every charge unit of an atom, creating an oscillating charge displacement and hence a dielectric polarization:

$$\boldsymbol{P} = \boldsymbol{D} - \varepsilon_0 \boldsymbol{E} = (\varepsilon - 1)\varepsilon_0 \boldsymbol{E}_0 \cos \omega t \qquad (27)$$

Equation (27) is a function of ν since the dipole formation does not always occur in phase with the applied field $\boldsymbol{E}$. Assuming that the oscillating dipole reacts quasi-elastically, vibrating with a characteristic frequency of $\nu_0 \approx (D/m)^{1/2}$ (D being the quasi-elastic force and m the mass of the oscillator) then the dispersion curve resembles a resonance curve. On increasing the applied frequency ν the phase displacement of the amplitude increases, and

approaches $\pi/2$ at $\nu = \nu_0$. At the point of maximum amplitude the induced current is in phase with the oscillating field (resonance), which results in a maximum of energy. Consequently, the dispersion curve for both ε and n suffers a sharp reversal in incline (abnormal dispersion), and for portions of the spectrum the refractive index increases with increasing wavelength. Since particular transmission directions in a crystal may absorb wave energy to a different extent, a direction dependence of coloring may be observed in polarized light known as pleochroism, and described in terms of, for example, X = weak, Y = strong, and Z = very strong or equivalently $X < Y < Z$.

The colors depend in a complex way also on their structural arrangements, types of chemical bonding, impurities, and defects. In this case an originally colorless transparent material may contain pigments and appear colored. Then, nonhomogeneities (precipitates or caverns) with dimensions of several hundreds of nanometers may generate internal scattering, leading to a colored luster (opalescence or labradorescence). The imperfections of crystals such as interstitial atoms or vacancies may cause local charge accumulations or lattice distortions which interact with light (color centers). The simplest type of point defect creating colors is the F center, which is a negative ion vacancy containing captured electrons. The most important origin for selective absorption is ion coloring, which is an intrinsic property of the crystal structure. The most effective chemical elements are cations of the transition metals (Ti, V, Cr, Mn, Fe, Co, Cu, and Zn) because they are able to occur in several oxidation states. The effect of absorption by the excitement of electrons in incompletely occupied d or f orbitals (d–f transitions) is called central ion coloring. The position of the absorption bands is determined by the coordination shell (symmetrical arrangement of the neighbor anions or charged complexes) which influence the splitting of the energy levels of the orbitals. Small variations in the interatomic distances, symmetry, or charges may have such strong effects on the colors of the crystal. For example, alumina (Al_2O_3), which is generally colorless, becomes ruby red if Cr^{3+} ($3d^3$ configuration) is added (allochromatic color), whereas the corresponding pure compound chromia (Cr_2O_3) exhibits a deep green color (idiochromatic color). Strong variations in color due to the distortion of the coordination polyhedra around the central ion (Jahn–Teller effect) have been found in, for example, Cu^{2+} compounds ($3d^4$ configuration), where the colors may change from green to deep blue, or in Mn^{3+} compounds ($3d^4$ configuration, pink and red colors).

Compared to central ion coloring, which covers electron transitions within the shell of a single ion, electron movement overlapping orbitals of adjacent ions may create an absorption of a 1000-fold intensity and result in nearly opaque materials. This effect is known as charge transfer coloration, and is observed between two cations, two anions, or between a cation and an anion. The absorption bands are rather broad, and spread from the ultraviolet to the visible region. The basic mechanism is a repetitive oxidation–reduction process between differently charged ions with overlapping d orbitals, as in the neighboring pairs Fe^{2+}–Fe^{3+}, Mn^{2+}–Mn^{3+} or Ti^{3+}–Ti^{4+}. Opportunities for such an arrangement occur preferentially by

coupled substitution of the constituents in the solid solution series, thus preserving the charge neutrality of the crystal structure.

If octahedra of differently charged central iron ions possess a common face, d-orbital overlapping permits oscillating electron transitions and thus charge transfer coloration. Generally, charge transfer is possible between octahedral coordination polyhedra with a common face or common edges (t_{2g} orbitals overlapping) and between tetrahedrally coordinated ions with common edges (e_g orbitals overlapping, common faces are generally impossible). The phenomenon of pleochroism, that is, direction-dependent selective absorption, is caused by a preferential unidirectional periodic arrangement of absorbing atoms as interconnected chains or clusters of coordination polyhedra. Thus, the particular orbitals permitting electron transitions between certain energy levels or between overlapping orbitals are in parallel and thus simultaneously excited by light waves of a certain oscillation direction.

An excellent tool for the calculation of energy level splitting and the width of absorption bands and hence for the prediction of color effects is the ligand field theory introduced by Bethe in 1929.

Extinction (intensity decrease of radiation passing through a medium) is a composite effect of scattering and absorption. If scattering is neglected, the resulting amplitude A of monochromatic light (λ_0) and initial amplitude A_0 after transmission through an absorbing medium of thickness d is

$$A = A_0 \exp(-2\pi k d/\lambda_0) \tag{28}$$

where k represents a characteristic material constant known as the absorption coefficient; this is Beer's law. It depends on the refractive index n according to

$$k = n\mu_a$$

yielding

$$n = \lambda_0/\lambda = k/\mu_a \tag{29}$$

where μ_a is the index of absorption and λ is the wavelength in the absorbing medium.

With $\boldsymbol{I} = A^2$ we obtain for the intensity

$$\boldsymbol{I} = A_0^2 \exp(4\pi k d/\lambda_0) \tag{30}$$

The term $4\pi k/\lambda_0$ is called the modulus of extinction. In transparent materials, k is usually very much less than 1 but it may be 3–5 for opaque substances. Like n, k is a rather complex function of λ. Equation (30) gives for the light vector $\boldsymbol{E}$ see Eq. (2)

$$E = E_0 \exp[t/T - d/\lambda_0(n - \mathrm{i}k)] \tag{31}$$

meaning that for absorbing media a complex index of refraction n_{refr} has to be inserted with

$$n_{refr} = n - \mathrm{i}k = n(1 - \mathrm{i}x) \tag{32}$$

Consequently, in the presence of absorption, the symmetry dependency of optical properties of matter can no longer be described using the relatively comprehensive indicatrix, except in the case of the cubic system where the complex vector of refraction n_{refr} equals nk. The generalization of the indicatrix for anisotropic systems gives a second-order surface with complex refractive indices as radial vectors [9]. With the complex units we obtain for the absorbing indicatrix

$$\frac{x_2}{n_{k1}^2} + \frac{y_2}{n_{k2}^2} + \frac{z_2}{n_{k3}^2} = 1 \tag{33}$$

which becomes, after a coordinate transformation, a system with the principal axes

x, y, and z:

$$\frac{x^2}{n_{k11}^2}+\frac{y^2}{n_{k22}^2}+\frac{z^2}{n_{k33}^2}+\frac{2xy}{n_{k12}^2}+\frac{2yz}{n_{k23}^2}+\frac{2xz}{n_{k13}^2}=1 \tag{34}$$

Similarly to nonabsorbing optically anisotropic matter, two wavefronts propagate through the absorbing crystal but, in contrast to the nonabsorbing material, they possess not only two different refractive indices but also two different absorption coefficients. Both waves are usually elliptically polarized and oscillate perpendicularly to one another with the same direction of rotation. The special condition of linearly polarized light exists only in optical symmetry planes.

In optically uniaxial systems the complex indicatrix consists of two surfaces for n and nk with a common revolution axis of different lengths. The surfaces are not ellipsoids, as in the case of nonabsorbing materials, but ovaloids. However, an ordinary and an extraordinary ray of linearly polarized light also exists.

Optical axes in the sense of quasi-isotropically behaving directions no longer exist in the orthorhombic, monoclinic, and triclinic systems. With increasing absorption the optical axes split, perpendicular to the particular optical symmetry plane, into two axes of preferential polarization status. These four axes are called the winding axes, because light propagating in these directions does not oscillate arbitrarily but is circularly polarized with an opposite direction of rotation. There are two principal directions in which only one refractive index is effective but two absorption indices, resulting in linearly polarized light and, alternatively, another two directions of equal absorption coefficients but different refractive indices.

1.2.5 Polarization

A wave is termed linearly polarized if its excursions take place in only one direction perpendicular to the direction of propagation. Most of the waves emitted from one source only are polarized. Electromagnetic waves emitted from one atom are linearly polarized as well. Waves resulting from the superposition of different primary sources are usually unpolarized, for example, the light of an ordinary light bulb is completely unpolarized, because it is emitted from oscillations from many different atoms.

Four effects can produce polarized light: absorption, reflection, scattering, and birefringence.

Polarization by Absorption

Polarization by absorption can be carried out using polarizing foils which consist of long chains of hydrocarbon molecules. These molecules must be oriented in one direction, which can be achieved by stretching during the fabrication process. Incident rays induce electrical currents, if $\boldsymbol{E}$ is parallel to the chains. Consequently, the energy of light is absorbed. If $\boldsymbol{E}$ is perpendicular to the chains, no absorption takes place, and the light is transmitted. Therefore, the direction perpendicular to the chains is called the transmission axis. This effect is used for polarizers and analyzers in polarization light microscopy.

Polarization by Reflection

Randomly polarized light reflected from polished flat surfaces is partially linearly polarized. The amount and kind of polarization depend on the angle of incidence,

the refractive indices, and the crystal symmetry of the reflecting material and the quality of the reflecting surface. Generally, the incident light impinging the surface at an angle ϕ_i is partially refracted at an angle of ϕ_{refr} and partially reflected by $\phi_{refl} = -\phi_i$ (see Fig. 5). In nonabsorbing, isotropic materials there is a critical angle of incidence ϕ_{ic} at which the reflected beam is nearly totally linearly polarized. ϕ_{ic} is known as Brewster's angle or the polarization angle.

At Brewster's angle, the following relationship is valid:

$$n = \frac{\sin \phi_i}{\sin \phi_{refr}} = \frac{\sin \phi_i}{\sin(90° - \phi_i)} = \tan \phi_{ic} \quad (35)$$

Moreover, the refracted beam is also partly linearly polarized, and oscillates perpendicularly to the reflected ray. Consequently, almost no reflection occurs for linearly polarized incident light (impinging with ϕ_{ic}) with an electric vector vibrating in the plane of incidence because the vibrations induced in the dielectric medium would then oscillate parallel to the virtually reflected ray. This fact would violate the rule that energy cannot be transmitted in the direction of oscillation.

In nonabsorbing anisotropic crystals, a plane-polarized incident beam is split into two perpendicularly vibrating components which are usually reflected to different extents (anisotropy of reflection). Both beams superimpose again, forming a linear oscillation with a resulting vibration direction which appears slightly rotated toward the direction of the stronger reflected beam. The reflected beam then exhibits a different azimuth of oscillation relative to the incident beam and a smaller intensity. In special orientations no variation in the polarization direction appears. This case is known as uniradial reflection, and gives maximum reflectivity. Unpolarized incident light consequently becomes partially polarized light due to the influence of the reflected polarized light with the highest intensity.

Polarization by Scattering

Polarization by scattering takes place if unpolarized light strikes dipoles such as molecules and if $\boldsymbol{E}$ is parallel to the dipole axis of the molecules. Under these conditions the electrical field of incident light causes the emission of polarized light from the oscillating dipole (i.e., the molecule) perpendicular to the propagation direction of the incident beam [4].

Polarization by Birefringence

The case where the incidence of a linear polarized wave on an optically anisotropic crystal (see Sec. 1.2.3 of this Chapter) under the condition that $\boldsymbol{E}$ and the optical axis form an angle of 45° is interesting. The ordinary and extraordinary beams are in phase at the beginning and have the same amplitude. After passing through a thickness of $\lambda/4$, the phase difference increases up to 90°. Thus, the x and y components of the electric field are $E_x = E_0 \sin \omega t$ and $E_y = E_0 \sin(\omega t + 90°) = E_0 \cos \omega t$, respectively (of course $E_z = 0$). $\boldsymbol{E}$ rotates on a circle, that is, the wave is circularly polarized. Analogously, under the same geometrical conditions, the wave after passing through a thickness of $\lambda/2$ exhibits a phase difference of 180°. In this case $\boldsymbol{E}$ became linearly polarized, and the components of $\boldsymbol{E}$ are $E_x = E_0 \sin \omega t$ and $E_y = E_0 \sin(\omega t + 180°) = -E_0 \sin \omega t$, respectively. The direction of polarization is twisted up to 90° in comparison to the incident beam.

1.3 Imaging

This Section deals with basic knowledge of the nonaberrated imaging of objects with systems of lenses (objectives) into a real and magnified image which can be viewed by other systems of lenses (oculars) or which can be projected onto planes for the exposure of films or television (TV) cameras (projectives). First of all, imaging should be treated in terms of scattering and interference of electromagnetic waves.

1.3.1 Imaging as a Scattering Phenomenon

Imaging as a scattering phenomenon has been dealt with systematically by E. Abbe. The aim is the definite correlation between the microscopic object and its scattering image which is generated in the exit pupil of the microscope. For the formation of an image which shows similarity to the original object, in principle the collection of all light scattered from the object is necessary. Since the aperture is limited to prevent light from entering the front lens of the objective, in reality only a part of the scattered light is able to contribute to the image. Therefore, the image is more or less dissimilar compared to the original object (Fig. 9), which means that the resolution remains below the theoretical limit. The borderline case is reached if only the zero and first scattering orders are able to enter the front lens. If only the zero order can enter the objective, no structure will be visible in the image. Therefore, the resolution limit d_c is given by

$$d_c = k\lambda(n \sin \sigma)^{-1} \tag{36}$$

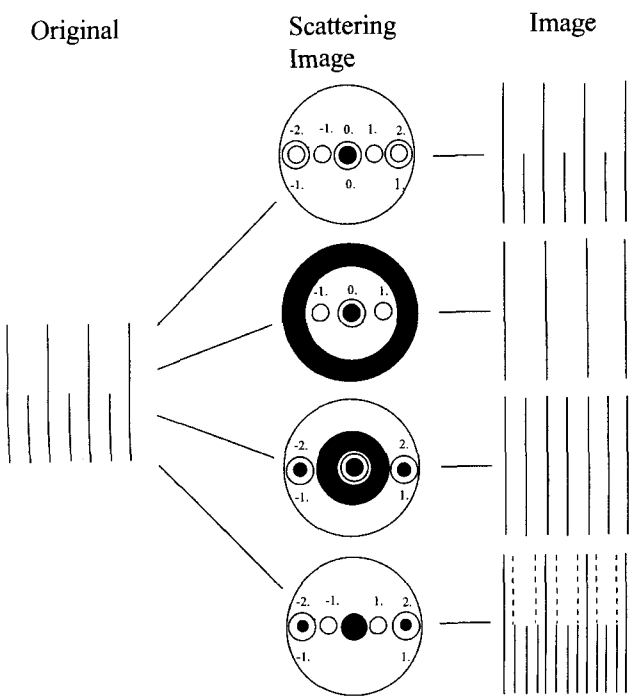


Figure 9. Correlation of the original object and its image (Abbe's experiments) for several contributing scattering orders [12]. 'Scattering image' indicates the intensity distribution in the back focal plane of the objective, which can be manipulated with the aperture diaphragm (positioned at the primary intermediate image—compare Fig. 11). (a) The image also reproduces the lines with half distances correctly, because for these lines too, the zero and the first scattering orders are included (for the lines with doubled distance even the zero, first, and second orders are included). Case (b) indicates the effect of the resolution limit (lines with half distances disappeared because their first scattering order is excluded). Cases (c) and (d) show the falsification by excluding certain orders of scattering.

where σ is the half angle of the most oblique light rays that enter the front lens and n is the refractive index of the medium between the object and the objective. The term $n \sin \sigma$ therefore characterizes the capability of collecting light, and is called numerical aperture A_n. Obviously, the resolution limit can be lowered by high angles of σ, by immersion substances with refractory indices $n > 1$ as well as by the application of shorter wavelengths. k is a constant which is 0.61 for luminescent points and 0.5 for opaque points if they are completely resolved and equal to 1 for an intensity decrease of 20% between the two maxima of emitted or scattered light.

1.3.2 Beam Paths

In the simplest case a microscope consists of two convergent lenses (Fig. 10). At the objective lens the essential beam paths for microscopical imaging are indicated:

- Beams which propagate parallel to the optical axis pass the back focal point (F'). More generally, all parallel beams, after passing the lens, intersect each other at one point in the focal plane.
- All beams which pass the lens through the optical axis in 0 (e.g. PP') are not changed in the direction of propagation.

Using the principal beam paths, the imaging of convergent lenses can be dealt with (Fig. 10). The lens equation describes the correlation between the focal distance f and the distances of the object d_o and of the image d'_o:

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d'_o} \tag{37}$$

The lateral amplification of the objective A_l and the magnification of the ocular M_{oc} are given by

$$A_l = d'_o/d_o = d'_o/f - 1$$
$$M_{oc} = 250/f \tag{38}$$

respectively.

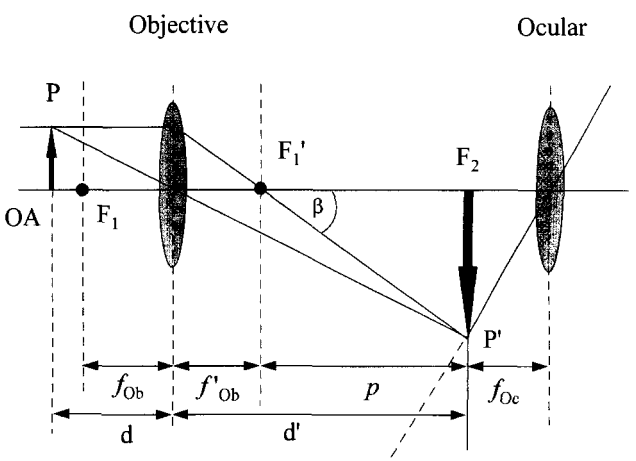


Figure 10. Schematic illustration of the simplest microscope, consisting of two convergent lenses, the objective and the ocular F_1 and F'_1 indicate the front and back focal points of the objective, that is, the interaction of its front and back focal planes with the optical axis OA, respectively. Beam paths indicate the generation of an inverted, real, and magnified image P' at a distance $d'_o > 2f'_{Ob}$ from an object P positioned at a distance $2f_{Ob} > d_o > f_{Ob}$, which is the common situation in the optical microscope. (If the object were positioned at $d_o = f_{Ob}$, the image would be generated at infinity. For $d_o > 2f_{Ob}$ the image of the object would be real and inverted but reduced in size. Finally, for $d_o < 2f_{Ob}$ the image would be virtually upright and magnified.) The magnified image is viewed by the *ocular* within the focal plane (f_{oc}) of this lens. Therefore, it cannot simply be projected virtually onto a film or a TV camera.

In Eq. (38) the magnification of the ocular M_i is expressed with respect to a conventional visual range of 250 mm. The total magnification of the optical system consists of the amplification of the objective as well as the magnification of the ocular:

$$M_t = A_1 M_{oc} \tag{39}$$

If the first (intermediate) image is not virtually projected from the focal plane of the ocular ($d_o'' = f_{oc}$) but really projected by a system, called *projective*, from a distance $2f_{oc} > d_o'' > f_{oc}$, a magnified real image is produced which can be documented using film or a TV camera. The total lateral amplification of such systems is

$$A_t = A_{obj} A_{pro} \tag{40}$$

Image Aberrations

Axial aberrations include both those designated as spherical and chromatic. Spherical aberration occurs with light of a single wavelength—different parallel rays pass the lens and are brought to a focus at different points along the optical axis, because they have passed through different zones of the lens. Even in well-corrected lenses there may be some residual effect, this being called zonal aberration. In chromatic aberration the various wavelengths of white light from a point lying on the axis are brought to a focus at different points along the axis, because in the application range of optical materials n increases with rising wavelength, that is, the blue light is more strongly refracted than the red (normal refractory dispersion). Even in well-corrected lenses there is usually a small secondary spectrum effect.

Off-axial aberrations include field curvature, coma, astigmatism, and distortion. Field curvature, which can also be due to astigmatism, occurs when the lens fails to produce a flat image of a flat object. Coma is the aberration that causes a point object to be imaged with a flare radial to the field. Astigmatism causes a point object to be brought to two focal points, one above and the other along the direction of the axis; also, a point is imaged as a line tangential to the field at one focus and radial to the field at the other focus.

In reflection microscopy, the flat field correction is often important. For studies with polarized reflected light, the importance of strain-free lenses is essential.

Depth of Sharp Focus

The depth of focus Δz, which is the distance between the positions of focal points at which the image appears with acceptable sharpness, can be estimated by the Rayleigh equation

$$\Delta z = \frac{\lambda}{n \sin^2 \sigma} \tag{41}$$

which may be combined with Eq. (36) to give a relationship between longitudinal and lateral aberration:

$$\Delta z = \frac{4n}{\lambda} \left(\frac{\lambda}{2n \sin \sigma} \right)^2 = \frac{n\lambda}{A_n^2} \tag{42}$$

Beam Path in a Composed Microscope

For scientific microscopes it is necessary to incorporate optical elements such as prisms, polarizers, and compensators, as well as illumination plane mirrors (in reflection microscopy) into the beam path between the objective and the ocular. This can be accomplished by using objectives with an infinite image distance. If the object is then positioned in the focal

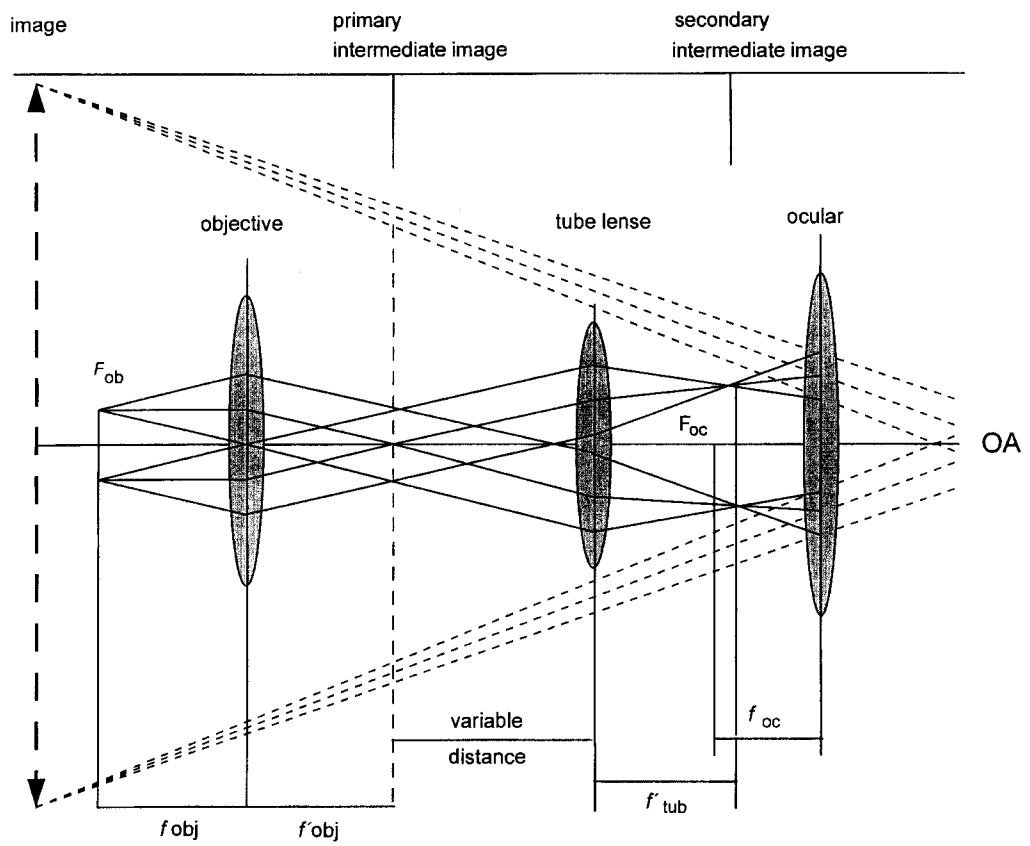


Figure 11. Beam path in a composed microscope with a tube lens to convert the intermediate image of the objective from infinity (note that the parallel beams of each object point between the objective and the tube lens) into a finite distance (*secondary intermediate image*). In the region of parallel beam paths between the objective and the tube lens, additional optical elements can be assembled without disturbing imaging.

plane, all beams emitted from one point of the object are parallel after passing the objective. In order to obtain an image at a finite distance, an intermediate lens is introduced, called the tube lens (Fig. 11). This lens produces a real image which can be magnified by an ocular or a projective, as already described. The tube lens is characterized by the tube factor q_∞. The total magnification for such a system can be written as

$$M_t = M_{obj} q_\infty M_{oc} \tag{43}$$

1.4 References

[1] R. Hooke, *Micrographia*, Royal Society, London **1665**, Reprint **1961**.

[2] H. Beyer in: *Handbuch der Mikroskopie*, 3rd ed. (Ed.: H. Riesenberg), Technik, Berlin **1988**, Chap. 1.

[3] H. Beyer, *Jenaer Rdsch.* **1973**, *18*, 159.

[4] P. A. Tipler, *Physics for Scientists and Engineers*, 3rd ed., extended version, Worth Publishers, New York **1991**, part V.

[5] J. Orear, *Physics*, Macmillan, New York **1979**, Chap. 26.

[6] V. F. Weisskopf, *Sci. Am.* **1968**, *9*, 60.

[7] M. Berek, *Central. Mineral. A* **1931**, *6*, 198.

[8] M. Berek, *Neues Jahrbuch für Mineralogie*, Abhandlungen Beilage 64, Abt. A **1993,** p. 123.
[9] M. Berek, *Fortsch. Mineral.* **1937,** *22*, 1.
[10] E. N. Cameron, *Ore Microscopy*, Wiley, New York **1961**.
[11] S. L. Swartz, *IEEE Trans. Electric Insulation* **1990,** *25*(5), 935.
[12] K. Michel, *Die Grundzüge der Theorie des Mikroskops in elementarer Darstellung*, 2nd ed., Wiss. Verlagsges., Stuttgart **1964**.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

2 Optical Contrasting of Microstructures

In the preceding sections the fundamentals of light and its interaction with matter were summarized. Thus, a variety of optical methods have been developed in order to characterize microstructures of natural substances, minerals, and artificial materials. These methods make definite use of various interactions between matter and light wave parameters, namely the change of amplitudes (i.e., intensities), variations in the wavelength distribution (i.e., the colors), and modifications of the state of polarization and phase shifts. The origin of these effects is based principally on the bonding and three-dimensional arrangement of the atoms (see Sec. 1.1 of this Chapter). Additionally, some methods in incident light microscopy use mainly indirect effects due to special preparation of the sample surfaces. In this case the surfaces are chemically and/or mechanically prepared to establish local excavation effects which may indicate variations of composition as well as of phase and defect structures. The resulting effects in roughness, surface waviness, etc., are transformed into local fluctuations of the reflected intensity which can be imaged with the bright or dark field technique, whereas a smooth waveform of the surface effects, by local phase, shifts in the reflected beam which can be converted into intensity effects by interference contrast methods. The following sections give an overview of the principles needed for a basic understanding rather than details of topical technical equiment, which can be obtained from the producers of optical microscopes. Therefore, descriptions of mechanical construction, the applied light sources, and other technical preconditions which are well developed in current systems and usually do not concern the operator, are also omitted in favor of a clear explanation of the optical principles for the various methods of imaging.

2.1 Transmission Microscopy

Light microscopy of transparent materials using a transmitted beam is a well-established domain of the geosciences and has become the most important tool for the geological and mineralogical characterization of rocks and minerals. The application of transmission light microscopy is also extended to include artificial materials, and is therefore used in the material science of glasses, polymers, and ceramics.

The principal arrangement of the optical elements is shown in Fig. 1 for the two important cases of imaging in transmission microscopy, indicating the orthoscopic and the conoscopic beam paths. However,

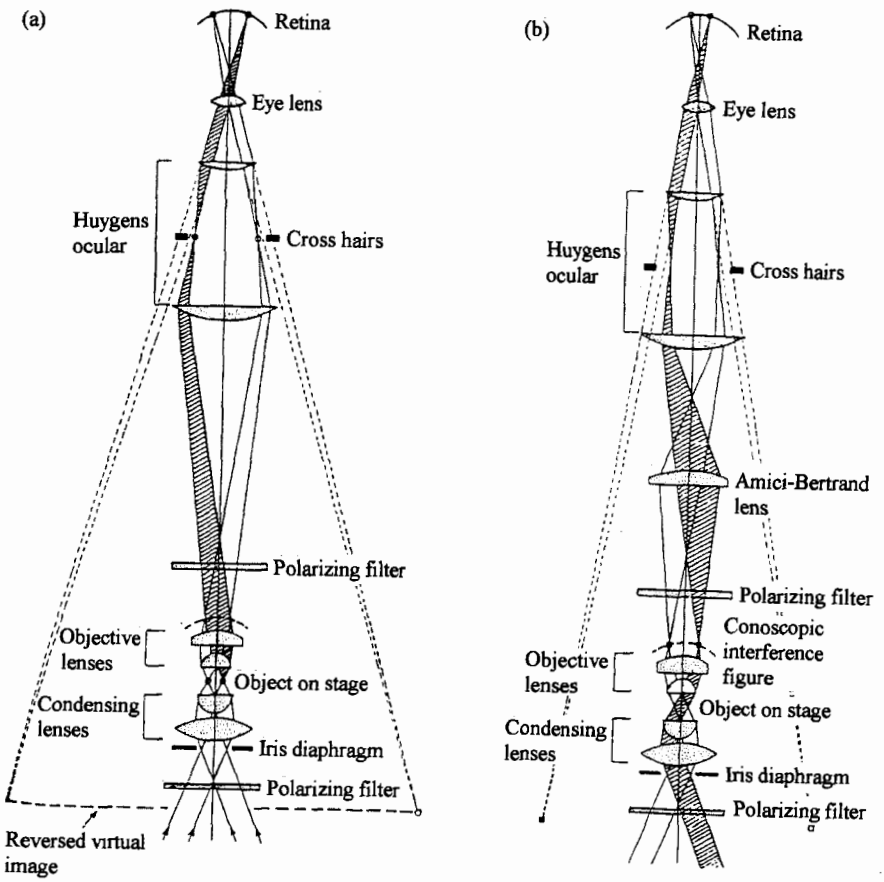


Figure 1. (a) Orthoscopic and (b) conoscopic imaging in transmission microscopy. (Adapted from Wahlstrom [1].)

several optical elements have to be eliminated or added for the different techniques of contrasting which will be discussed in the following sections.

2.1.1 Bright Field

In transmission as well as reflection microscopy, bright field illumination is the most generally applied basic technique. An efficient and homogeneous principle of illumination was developed by A. Köhler in 1893 [2]. The light passes the sample where it interacts with scattering objects (see Sec. 1.2 of this Chapter). The illumination is designed in such a way that the zero order of scattered light can also enter the objective and therefore contribute to imaging (see Sec. 1.3.1 of this Chapter). Because the image is generated with non-polarized light, for bright field microscopy no polarizers are used (compare Fig. 1).

Bright field illumination is used to recognize colors and interfaces of translucent

materials as well as to distinguish between translucent and opaque materials. However, most of the nonmetallic samples exhibit near transparency in the wavelength range of visible light. Therefore, several coloring techniques are used for imaging in bright field [3, 4]. Interfaces become visible due to refraction and reflection effects caused by the mismatch in refractive indices between two adjacent materials. Cracks in glasses and ceramics or minerals can easily be recognized, and also material of the same composition may reveal clearly visible grain boundaries due to orientation effects.

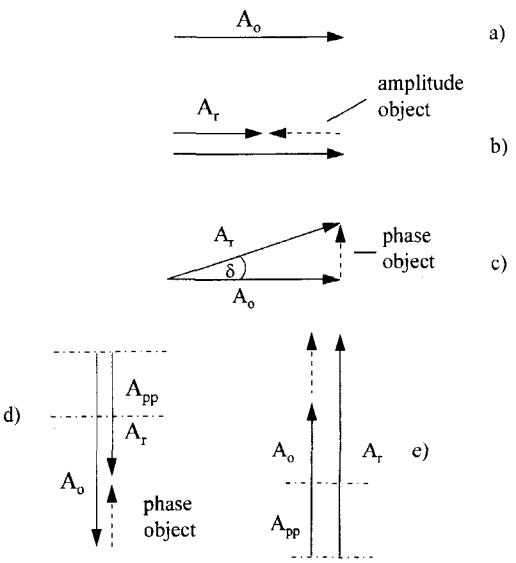


Figure 2. Vectorial representation of phase-related for (a) an incident wave, (b) the wave resulting after scattering at an amplitude object, (c) the wave after scattering at a phase object, and the transformation into an amplitude effect with the contrast technique of (d) negative or (e) positive phase contrast (A_0 is the amplitude of the incident wave, A_{pp} is the amplitude of the wave after passing the phase plate, and A_r is the resulting amplitude of the wave).

2.1.2 Dark Field

Contrary to bright field microscopy, the scattered light waves of zero order (i.e., the direct light) cannot generally contribute to the image which is generated by at least two other orders of scattered light (see Sec. 1.3.1 of this Chapter). The contrast in dark field microscopy is also established by reflection and refraction effects at interfaces in between the sample, where the (local) direct light is also finally reflected into the objective and produces bright spots of direct light (i.e., light waves of zero order) in the image.

2.1.3 Phase Contrast

Besides the amplitude objectives, which convert the amplitude of the scattered light and therefore produce a direct intensity contrast (Fig. 2b), phase objects also exist, which have negligible effects on the amplitude but shift the phase position of the wave in a different way than the adjacent microstructure of the sample (Fig. 2c). For such samples, the problem consists of transforming such phase differences into a difference in the amplitudes which become visible as intensity effects. In 1932, F. Zernike proposed the phase contrast technique as a consequent application of Abbe's theory of image formation on phase objects [5–7].

This method can be summarized using the schematic sketch in Fig. 3 as well as the vectorial diagrams in Fig. 2. A parallel beam passes the plane of the sample (Fig. 3, position 1) where a phase object (compare Fig. 2c) is positioned. The parallel incident light (zero order) is focused onto the optical axis in the back focal plane (Fig. 3,

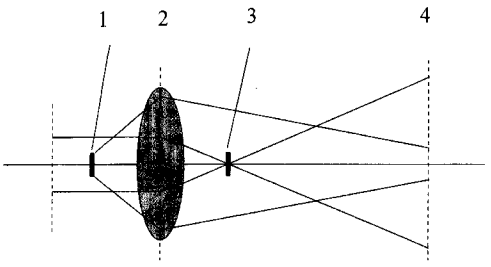


Figure 3. Schematic representation of the phase contrast technique: 1, sample; 2, objective; 3, phase plate positioned at the back focal plane of the objective; 4, first image of the object. For an explanation, see the text.

position 3) of the objective (Fig. 3, position 2), whereas the higher orders of scattered light are distributed in the back focal plane. Therefore, the zero-order light can be manipulated separately within the back focal point of the objective. Therefore, at this point (Fig. 3, position 3) a phase plate is positioned which carries out a defined phase shift (negative or positive) to the zero-order light. Consequently, in position 4 of Fig. 3 the waves of zero order are shifted relative to the nonshifted waves of the scattering waves of higher orders. Thus, the initial phase shift of negligible amplitude effect at position 1 is transformed into amplitude effects of negative (see Fig. 2d) or positive (see Fig. 2e) phase contrast. With lower transparency of the phase plate the intensity of the zero-order light can be reduced, which leads to a stronger intensity effect in this contrast technique. Important applications in chemistry and mineralogy of phase contrast techniques are the determination of refractory indices (immersion method [8]) of particles as well as the morphological characterization of particles with different refraction (e.g., see Correns [9]). Other scientists have investigated the segregation of peristerite–plagioclas, applying dark field as well as phase contrast techniques [10].

The phase contrast technique represents contrasting by an absolute scale of phase shifts. Today the most important fields of application are in biology and medicine, that is, the investigation of organic matter. In many investigations on inorganic substances, contrasting with differential techniques is preferred, such as the differential interference contrast (see Sec. 2.1.5 of this Chapter).

2.1.4 Polarization Microscopy

A typical instrumentation for polarization microscopy is shown in Fig. 1. In contrast to applications in incident light microscopy the images obtained by polarized light in transmission microscopy can be evaluated more quantitatively since the effects observed are much easier to measure and understand.

In transmission microscopy it is easy to obtain information on the crystal system (cubic or tetragonal, trigonal, or hexagonal (uniaxial) or orthorhombic, monoclinic, or triclinic (biaxial)) as well as twinning, orientation, texture, and stresses. Isotropic cubic or amorphous materials can be distinguished from anisotropic materials by using crossed polarizers because of the permanent darkness at all orientations to the transmitted beam. If the transmitted beam propagates parallel to an optical axis, anisotropic materials show distinct directions of isotropy or quasi-isotropy. Splitting of the propagating waves takes place if the transmitted beam is not parallel to an optical axis. Then there exist different velocities and

polarization modes (double refraction and birefringence; see Sec. 2.2.5 of this Chapter) now vibrating in perpendicular polarization planes. Both rays interfere after passing the analyzer and superimpose with the path difference Γ and the vibration directions of the wave vectors. If only the horizontal vibrating component of the resulting ray is transmitted, plane-polarized light with a wavelength and amplitude different to that entering the crystal is generated.

Thus, the dependency of the observed wavelength λ and intensity $\boldsymbol{I}$ can be explained as functions of (1) the wavelength of the incident beam, (2) the optical properties of the crystal size and symmetry of the indicatrix (extent of birefringence) and coefficients of absorption, (3) the relative orientation of the crystal (indicatrix) with respect to polarizer and analyzer, and (4) the thickness of the sample.

The interaction of plane-polarized light with the crystal depends on the particular cross-section of the indicatrix, which is normal to the transmitted beam. Assuming parallel incident light, this arbitrary cross-section is an ellipse with principal axes $n_{\alpha'}$ and $n_{\gamma'}$—neglecting the absorption and dispersion. This means that this cross-cut of the crystal exhibits a partial birefringence Δ' of

$$\Delta' = n_{\gamma'} - n_{\alpha'} \tag{1}$$

resulting in two separately propagating rays of velocity

$$v_1 = c/n_{\alpha'} \qquad v_2 = c/n_{\gamma'} \tag{2}$$

The time for crossing a crystal plate of thickness d is given by

$$t_1 = d\frac{n_{\alpha'}}{c} \qquad t_2 = d\frac{n_{\gamma'}}{c} \tag{3}$$

Therefore Γ is a function of the thickness d:

$$\Gamma = (t_2 - t_1)c = (n_{\gamma'} - n_{\alpha'})d = d\Delta' \tag{4}$$

where the units for r and d are nanometers and millimeters, respectively.

With an arbitrary orientation of the principal axes of the indicatrix section relative to the polarizer and analyzer, the particular polarization mode of the waves leaving the crystal, as well as the intensity of the observed beam behind the analyzer, can be obtained by splitting the incident wave vector $\boldsymbol{k}_{\text{res}}$ into the vectors $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$ within the crystal vibrating along the principal axes:

$$k_1 = k_{\text{res}} \cos\phi \qquad k_2 = k_{\text{res}} \sin\phi \tag{5}$$

where ϕ is the angle between $\boldsymbol{k}_{\text{res}}$ and $\boldsymbol{k}_1$ or $\boldsymbol{k}_2$. It has to be considered that the waves can be described physically as elliptically, circularly, or plane polarized. A more detailed study on the vector analysis of the polarization modes is given by Wahlstrom [1].

For coherent waves, the interference after the analyzer can be described, if the projection of the components onto the analyzer plane is considered:

$$\begin{aligned} k_1^{\text{A}} &= k_1 \sin\phi = k_{\text{res}} \sin\phi\cos\phi \\ k_2^{\text{A}} &= k_2 \cos\phi = -k_{\text{res}} \sin\phi\cos\phi \end{aligned} \tag{6}$$

If the vibration planes of the polarizer and analyzer are mutually perpendicular, $\boldsymbol{k}_1^{\text{A}}$ and $\boldsymbol{k}_2^{\text{A}}$ are of the same length but of opposite direction, and complete extinction would occur at the analyzer. However, according to Eqs. (9, Chapter 1) and (4) the phase difference is given by

$$\delta = \frac{2\pi}{\lambda}\Gamma = \frac{2\pi}{\lambda}(n_{\gamma'} - n_{\alpha'})d \tag{7}$$

and inserting for the intensity $I_i = |k_i|^2_{\max}$ gives

$$I^A = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\delta \qquad (8)$$

The intensity thus depends on ϕ and δ, resulting in extinction for $\sin(\delta/2) = 0$ and for $\phi = 0$ or $\phi = \pi/2$, that is, if the vibration planes of the resulting waves behind the crystal are parallel to the polarizer and the analyzer. A maximum intensity is obtained at a position of $\sin^2 2\phi = 1$ and $\phi = \pi/4 = 45°$. Hence, at the extinction position there is no beam splitting, but the unaffected transmitted beam is completely blocked by the analyzer.

Allowing the observation of an intensity at $\phi \neq 45°$ and varying Γ or δ, respectively, there is the extinction of particular wavelengths λ at $\Gamma = m\lambda$ for $\delta = 2m\pi$ (with $m = 0, 1, 2, \ldots$) and a maximum intensity of that wavelength for $\Gamma = (2m+1)\lambda/2$ for $\delta = (2m+1)\pi$, referring to destructive and constructive interference.

Characterization and Identification of Anisotropic Crystals

The above-mentioned effects are very useful for the characterization and identification of anisotropic materials (see Figs. 4, 5, 6 and 7). The four positions of extinction on rotating the crystal on the stage about the incident beam determine the exact location of the axes of the indicatrix. Its orientation relative to the crystallographic axes can be determined if the crystal exhibits clear morphological features such as facets, edges, twins, cleavage planes, epitaxial inclusions, or coatings. For systems of high symmetry, at least one of these characteristics should be parallel to the principal axes of the indicatrix (straight extinction). The case where only pyramidal facets are visible in these systems is called symmetric extinction. In monoclinic and triclinic systems the indicatrix may be inclined relative to the crystal axes (oblique extinction). The angle between crystal axes (a, b, or c) and an indicatrix axis (X, Y, or Z) is known as the angle of extinction. Hence, the angle of extinction can be used for the identification of members of solid solution series (e.g., feldspars and pyroxenes), because the orientation of the

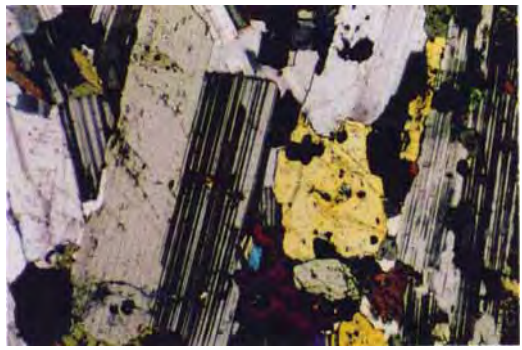


Figure 4. Thin section for petrological microanalysis. Monzodiorite, Gröba/Saxonia (feldspar, pyroxene). Magnification, 50×; thickness, 30 µm; crossed nicols. (Courtesy of M. Magnus, TU Mining Academy Freiberg, FRG.)

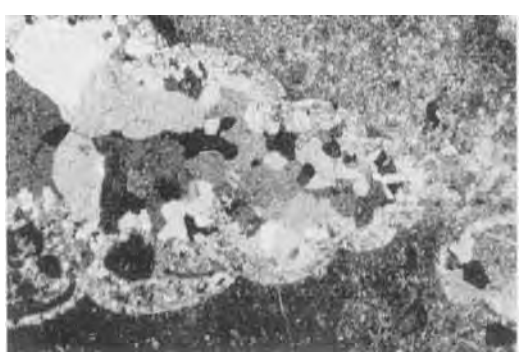


Figure 5. Thin section for paleontological analysis. Longitudinal section from a freshwater gastropod (*Anthracopupa*), Zwickau/Saxonia (consists of carbonatite grains). Magnification, 50×; thickness, 45 µm; parallel nicols. (Courtesy of M. Magnus, TU Mining Academy Freiberg, FRG.)

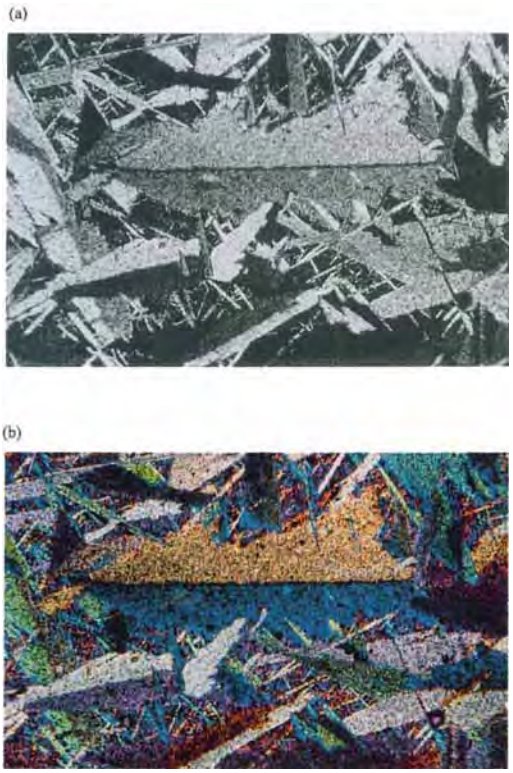


Figure 6. Thin section of a structural material (SiC): (a) without $\lambda/2$ plate and (b) with $\lambda/2$ plate. Polarized light; magnification, 200×. (Courtesy of U. Täffner and R. Telle, Max-Planck-Institute for Metals Research, Stuttgart, FRG.)

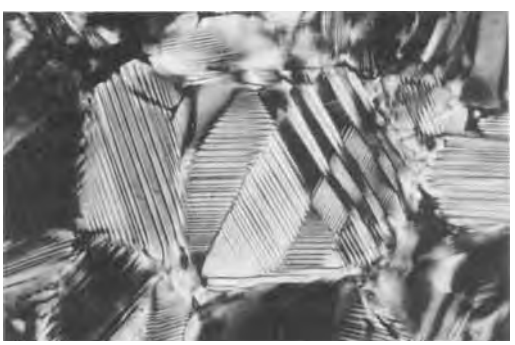


Figure 7. Thin section of a functional material (piezoelectrical $BaTiO_3$) showing the domain structure. Polarized light; magnification, 1000×. (Courtesy of U. Täffner and H. Rostek, Max-Planck-Institute for Metals Research, Stuttgart, FRG.)

indicatrix depends on the chemical compositions. Owing to eutectic growth, the materials can consist of several zones of various extinction angles.

Some crystals show diffuse extinction of particular segments called undulating extinction. If the crystal exhibits dispersion of the optical axes there is no complete extinction at one crystal position but only for a certain wavelength. As a result, the complementary colors become visible upon rotating the crystal, giving a bluish gray or dirty brown impression known as anomal extinction.

Equation (4) can be utilized for measuring thicknesses when the optical constants and the orientation of the crystal are known (e.g., Δ' is given or, for the recognition of unknown phases, the thickness is known). As Γ results in extinction at a particular wavelength λ at $\Gamma = m\lambda$ and a maximum intensity at $\Gamma = (2m + 1)\lambda/2$, at a given thickness d certain wavelengths from the incident white light are extinguished or weakened. The resulting complementary colors, known as interference colors, are a sensitive function of the thickness and birefringence, which are listed in the so-called Michel–Lévy chart, which is a color table correlating path differences in nanometers, thicknesses in millimeters or micrometers, and extents of birefringence. This is also the reason why the thickness of a sample in geoscience is normalized to 20–30 μm, which makes the identification of minerals much easier (therefore quartz with its gray color of the first order can be easily used as a standard in rocks).

Since the superposition of sinusoidal intensity fluctuations for distinct wavelengths yields a periodic change of interference colors with a repeatedly appearing

purple-red coloration for steps of 551 nm path difference ('red of first order'), the succession of interference colors may be arranged in several orders. The first order includes $\Gamma = 0$ (in this case the visible colors are black, dark gray (40 nm), lavender gray (97 nm), greenish white (234 nm), yellow (306 nm), orange (470 nm), red (536 nm), and purple red (551 nm)). The second order, consisting of violet, indigo, blue, green, yellow, orange, and red, has basically the same sequence but is clearer and brighter, whereas at higher orders the color separation becomes increasingly weaker due to the irregular mixing of wavelengths, path differences, and luminosities. Finally, the white light of higher order is visible. The interference colors may be used for the identification of chemical compounds and their orientation in multiphase materials.

The thickness of sections has to be adapted for the information required, remembering that with decreasing thickness the intensity and luminosity of interference colors also decreases. As an aside, many possibilities for the determination of optical constants and crystal orientations from the correlation between path difference and birefringence have been derived, and many optical accessories have been developed for quantitative measurements [11].

All optical accessories are inserted in the tube at a slot provided in a 45° position relative to the polarizer and the analyzer in order to gain maximum intensity (see Eq. (8)). The compensators add or subtract a known path difference to that caused by the birefringence of the sample, and hence generate variations of the interference colors. An increasing path difference (addition) is accomplished if the sample crystal and the compensator crystal are oriented in such a way that the vibration planes of both rays of the sample and the compensator are parallel and have an effect similar to that for a thicker sample, whereas the opposite orientation diminishes the path difference (subtraction, compensation, or retardation). To understand the observed phenomena quantitatively, the vibration direction of n_γ and the path difference Γ are marked on the compensator. The most common compensator is a special cut of a gypsum crystal, creating a path difference of 551 nm. This color is extremely sensitive to slight variations in wavelength, changing immediately to blue or other colors of higher order if a path difference is added, or to yellow or gray if a path difference is retarded. The gypsum plate can be used for the determination of the optical character by examination of the character of elongation l or of conoscopy. The character of elongation is attributed to crystallites which possess a preferred growth direction. Rotating the significant growth direction to a 45° position and inserting the compensator, blue interference colors indicate a positive character (addition) and yellowish colors a negative character of elongation (subtraction). In the case of uniaxial crystal systems growing in the direction of the c axis, the character of elongation is equal to the optical character. In other systems the character of elongation can be correlated to the optical character only if the orientation of the indicatrix relative to the crystallographic axes is known.

Another accessory for determining the optical character and fast or slow transmission directions in a sample is the quarter-wave plate ($\lambda/4$ plate), which

gives a path difference of $\lambda/4$ for a certain wavelength, usually for sodium light. Monochromatic elliptically polarized light is thus converted into plane-polarized light. It can be used for the very sensitive compensation of all wavelengths, except for sodium light, as well as for the quantitative measurement of very small path differences if it is combined with a precisely rotating analyzer (the Senarmont method [8]). Furthermore, the polarization mode of optically active crystals can be determined. In a similar way, λ plates are crystal plates which are able to extinguish a particular λ from the white light according to Eq. (7), and are used as very sensitive filters. If the preferred growth direction of a particular phase is known, compensation with a gypsum crystal may be easily used for texture analysis of ceramic parts. Parallel crystallites forming areas of the same interference colors reveal characteristic microstructural features introduced by the fabrication technique [12].

The use of a universal stage makes three-dimensional analysis possible, which can be plotted as a pole diagram.

Other types of compensators create a variable but exactly calibrated path difference. A widely used auxiliary instrument of this kind is the quartz wedge (the Soleil compensator), which yields a thickness-dependent change in Γ. The disadvantage is that very low path differences cannot readily be observed because of the difficulty of preparing a thin edge. The problem was solved by the Soleil–Babinet compensator, which uses two displaceable quartz wedges of different crystal orientation which yield a zero line due to subtractive interference. The Berek compensator generates a variable path difference by being tiltable around a principal axis of the indicatrix. Thus, path differences may be adjusted due to variations in thickness relative to the transmitted beam.

Conoscopy

Observation of virtual images of anisotropic crystals using a convergent beam at crossed polars, also known as interference figure microscopy (conoscopy, see Fig. 1), provides information on the number of optical axes, the optical axis angle, and finally the optical character of a crystal. Inserting the condenser, focusing the object at a high magnification (1000×, large aperture angle), and inserting the Amici–Bertrand lens produces an interference figure which results from the incident light forming a cone above the condenser lens. The clearest and most informative figures are obtained in uniaxial systems if the crystal is observed almost parallel to the optical axis, or in biaxial systems if observed in parallel to the acute bisectrix. These conoscopic figures consist of two systems of extinction patterns called isogyres and isochromates. In monochromatic light, both systems are black, whereas in white light the isochromates show the same color sequences as a quartz wedge. The isogyres refer to areas of equal vibration directions being parallel to the polarizer and the analyzer, and hence being extinguished. Since the crystal is observed simultaneously from many directions, depending on the aperture angle ϑ the transmitted beams propagate over different distances, causing a particular path difference Γ_ϑ:

$$\Gamma_\vartheta = n_\vartheta \lambda = \frac{d\Delta_\vartheta}{\cos\vartheta} \tag{9}$$

where n_ϑ is the number of observed isochromates and Δ_ϑ is the particular birefringence of the transmission direction ϑ. The isochromates thus represent figures of constant path differences Γ_ϑ which appear bright in monochromatic light at $\Gamma_\vartheta = \lambda/2, 3\lambda/2, 5\lambda/2, \ldots$ and black at $\Gamma_\vartheta = 0, \lambda, 2\lambda, 3\lambda, \ldots$ For uniaxial systems, a view parallel to the optical axis shows a cross-shaped isogyre with the branches parallel to the polarizer and the analyzer. This can be explained by considering the vibration directions of an ordinary beam o oscillating tangentially and an extraordinary beam e oscillating radially with reference to the optical axis figure. The isochromates form concentric rings of narrower interspacings and weaker intensities with increasing distance from the center. The center of the figure, also referred to as the melatope, can be interpreted as the position of the optical axis. If the optical axis is exactly adjusted to the axis of the microscope, the interference figure does not change upon rotating the stage since the uniaxial indicatrix is an ellipsoid of revolution (see Fig. 7 of Chapter 1). A slight mismatch of the orientation will cause modification of the cross until, at coarser misorientations, only single branches propagate through the eye field.

Biaxial systems exhibit hyperbolic curves as isogyres if the sample is observed parallel to the acute bisectrix. The vertices of the hyperbolas again represent the intersection of the optical axes, and their distance corresponds to the optical axis angle $2V_z$. Starting from a uniaxial conoscopic figure, the formation of distinct curves in the biaxial system can be explained by a separation of pairs of adjacent branches of the cross due to the splitting of the axis. Biaxial crystals with a very small angle of $2V_z$ exhibit a conoscopic figure similar to that of a uniaxial crystal. The isochromates form so-called Cassinian curves with minor symmetry with respect to the optical normal plane. On rotating a well-centered acute bisectrix, the hyperbolas join, forming a cross with branches parallel to the polarizer and analyzer if the optical axes plane is also parallel to the polars, and they swing out to the largest separation distance if the optical axes plane is oriented at 45°. With increasingly decentralized orientation, only single hyperbolic curves are visible with curvatures dependent on the optical axis angle. If the transmitted light propagates parallel to the obtuse bisectrix, diffuse isogyre branches may propagate through the eye field upon rotating the stage, but derivations from these figures can be made only with considerable experience.

For the quantitative explanation of conoscopic figures a model was developed by Becke (1905 and 1909) called the Skiodrome method, which considers the three-dimensional orientation of the vibration planes of the rays propagating through a crystal as a function of the angle of incidence and with respect to the position of crossed polars [13].

The optical axis figures are very useful for the identification of transparent matter by (1) determination of the symmetry of a crystal, (2) definition of the orientation of the indicatrix with respect to the crystal axes, (3) measurement of $2V_z$ of the acute bisectrix, (4) determination of dispersion, and (5) determination of the optical character.

In uniaxial systems the orientation of the c axis (optical axis) can easily be characterized. In monoclinic crystals the

optical normal yields the direction of the crystallographic b axis. The investigation of large areas using conoscopy provides information on texture which can be quantified if the orientations of the optical axis or the acute bisectrix are transferred to a pole diagram. A universal stage, rotatable in all space directions, may be a useful accessory for this task.

Considering the refraction of light emerging from the crystal, the optical axis angle $2V$ can be calculated from the scalar distance d between the melatopes if the optical axis plane is oriented in a 45° position relative to the analyzer and the polarizer. This is of particular importance for the determination of the members of a solid solution series because $2V$ is a sensitive measure of chemical concentration differences (compare the angle of extinction). Applying Snellius's law (see Eq. (21)) to the geometrical situation, we obtain

$$\phi_i = V_{obs} \qquad \phi_{refr} = V$$

and thus

$$\sin V = \frac{\sin V_{obs}}{n_\gamma} \tag{10}$$

The limit for an observed angle $2V_{obs}$ is given by the numerical aperture of the objective lens, which should thus be large. Measuring the scalar distance d between the melatopes using a ruler ocular, V_{obs} in air is calculated by

$$\sin V_{obs} = \frac{d}{2M}$$

or

$$\sin V = \frac{d}{2Mn_\gamma} \tag{11}$$

with M being Mallard's constant, which has to be ascertained for a particular microscope by measuring V_{obs} on a crystal with known V and n_γ according to Eq. (11). Formulas for the construction of tables or charts containing correlations between $2V$ and $2V_{obs}$ have been published [14, 15]. Another method for the determination of $2V$ makes use of the curvature of the isogyres in the 45° position. If the isogyre remains a straight line in all positions on rotation, then $2V$ equals 90°. As the curvature angle of the isogyre approaches 90°, the optical angle approaches uniaxiality. The angles in between may be calculated for the particular refractive indices and numerical apertures. Optical axis figures can also be used for the determination of strain-induced deformation of crystal structures through knowledge of all unconstrained constants. The method may be preferentially applied to uniaxial systems since any distortion of the structure at an oblique angle to the c axis results in biaxiality. The amount of axis splitting may be used to calculate the applied stress.

For this procedure the elasto-optical tensor must be known, for example cubic and thus isotropic magnesium-stabilized zirconia is constrained by MgO inclusions producing an interference color of a higher order in the vicinity of those particles [16]. Note also the deformation lamellae (light diagonal lines) and the fine perpendicular network of striations due to epitaxial tetragonal zirconia precipitates.

Remembering the fact that isogyres are areas of extinction of certain waves vibrating parallel to the polarizer and the analyzer, and taking into account that the melatopes may change their position as a function of λ, it is obvious that in white light illumination, dispersion in biaxial systems must reveal isogyre fringes of complementary colors.

Recognition of the optical character by conoscopy makes use of a gypsum first-order red plate inserted in a 45° position. The method is based on the retardation and enhancement of optical path differences producing a very sensitive change in coloration for the determination of the character of elongation.

Light emerging at any point of a crystal in a conoscopic view now consists of two components, the extraordinary beam and the ordinary component. For the exact determination of a crystal it is very important to know which quadrant is actually visible, especially in the case where the conoscopic figure is drastically off-center.

2.1.5 Differential Interference Contrast

Differential interference contrast in transmission microscopy is also a technique used to visualize phase objects. Since it is especially sensitive to small phase shifts, it is increasingly preferred to the phase contrast method. It operates with a lateral splitting of the image close to the resolution limit of the objective (shearing technique). The instrumentation is based on polarized light (see Fig. 1) and completed with Wollaston prisms for beam splitting as well as for beam junction. Technical solutions have been proposed by Francon [17] and also by Smith and Nomarski [18, 19]. The most common instrumentation is that proposed by Smith and Nomarski, and its principles are now explained (Fig. 8).

The primary beam is polarized (1) and impinges the first Wollaston prism (2) with an orientation of 45° to the optical axis of the prism, so that the polarized light is split into an ordinary and extraordinary beam of the same intensity. Both beams are focused on the focal plane (3) of the condensor lens (4) and consequently pass the object plane (5) in parallel but laterally shifted. The upper part of the optical scheme is designed rather symmetrically so that both beams can interfere after passing the second Wollaston prism (8) and the analyzer (9).

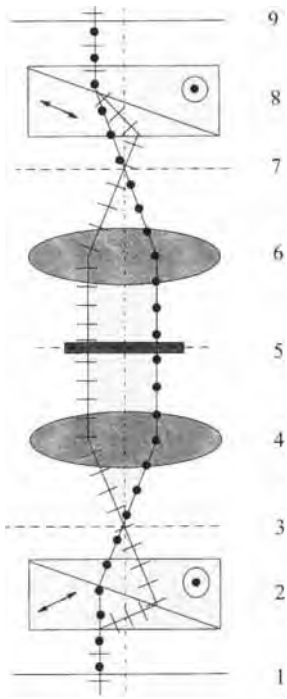


Figure 8. Optical scheme of the differential interference contrast technique proposed by Nomarski. 1, polarizer; 2, 8, Wollaston prisms; 3, focal plane of the condensor; 4, condensor; 5, object plane; 6, objective; 7, back focal plane of the objective; 9, analyzer.

Due to local phase objects, the beams are additionally phase shifted to each other because a difference in the effective wave paths has occurred. Thus, the resulting phase differences are transformed into

an amplitude difference, that is, into an intensity effect after final interference. Of course, a contrast appears only at microscopical positions with gradients of phase shifts and only in the direction of lateral beam splitting, which gives the images some impressive color and shadow effects. However, one must be more careful than in the phase contrast technique concerning the correct interpretation of the images.

2.2 Reflection Microscopy

Reflection microscopy is also a basic technique for the characterization of opaque substances. The condition is that the sample must reflect a significant part of the incident light in order to generate an image. Since the sample surface quality directly affects the image quality, sufficient surface preparation is necessary (see Sec. 2.3 of this Chapter). Generally, the sample surface should be positioned perpendicularly to the optical axis of the microscope, which can easily be achieved by applying the inverted microscope according to Le Chatelier's principle with the sample lying on a holder above the objective. This ensures optically correct positioning even for an irregularly shaped sample, as long as one flat surface is available.

The illumination is commonly inserted between the tube lens and the objective lens usually by means of semitransparent mirrors or—for quantitative investigations with polarized light—by totally reflecting prisms. Therefore, the objective lens additionally acts as a condenser unit (Fig. 9).

2.2.1 Bright Field

Bright field microscopy is characterized by imaging including the zero order, that is, of direct light (see Sec. 1.3.1 of this Chapter). In reflected bright field illuminations, light that strikes flat features in the image (perpendicular to the optical axis) is reflected into the objective and appears

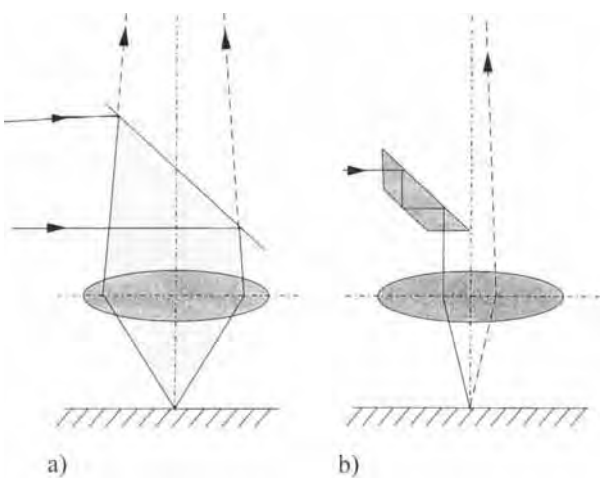


Figure 9. Schematic illustration of the illumination for reflection microscopy in bright field modes indicating the application of (a) a semireflecting mirror and (b) a Berek prism. For quantitative measurements with polarized light, prisms are used instead of the semireflecting mirror.

bright in the image. Contrast is achieved if significant variations of absorption and reflection of the flat features occur (see Sec. 1.2.1 of this Chapter). Additionally, light that strikes inclined features is scattered and appears dark in the image. These local excavation effects are indirect signals for the microstructure. They are often produced by certain chemical, mechanical, or other surface preparation techniques in order to recover phases, grains or twin boundaries, inclusions, dislocation etch pits, etc. (see Sec. 2.2 of this Chapter).

2.2.2 Dark Field

Dark field illumination often produces to some extent a reversal of the image contrast as compared to bright field microscopy. The light that is scattered is collected and appears bright in the image, while the light which reflects directly (i.e., of zero order; see Sec. 1.3.1 of this Chapter) does not contribute to a flat surface. However, for local, inclined surfaces, the zero order of reflection can be allowed to contribute to the image and produce bright spots. This technique is therefore very sensitive to surface irregularities and features such as grain or twin boundaries, slip lines, etch pits, and scratches. Because bleeding white light of the zero order of reflection is excluded, color effects of the microstructure may be visible, which often cannot be observed in bright field illumination.

The illumination for dark field reflection microscopy is carried out using a ring-shaped mirror (Fig. 10 (2)) which surrounds the imaging beam path and replaces the semitransparent mirror. The light is reflected by the parabolic mirror with oblique incidence onto the object surface (Fig. 10 (4)).

2.2.3 Phase Contrast

The technique of phase contrast has been described in some detail in Sec. 2.1.3 of this Chapter for the case of transmission microscopy. The theory of reflecting

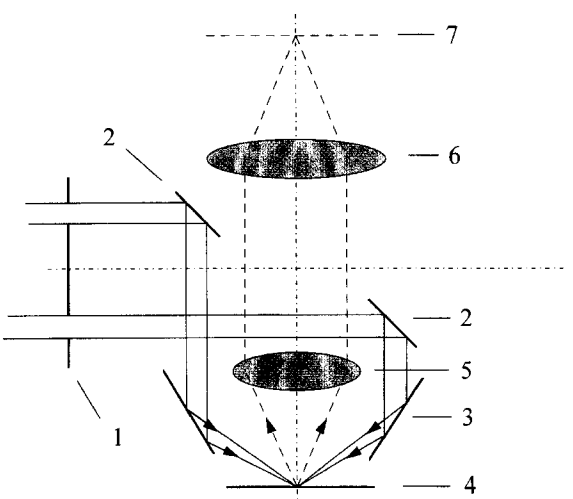


Figure 10. Schematic illustration of dark field imaging with all-round inclined illumination. 1, dark field diaphragm; 2, ring-shaped mirror; 3, parabolic mirror; 4, object; 5, objective; 6, tube lens; 7, back focal plane of tube lens.

microscopy is similar to that of transmission microscopy. Today, this method has been almost completely replaced for most applications by the technique of differential interference contrast, mainly due to its higher sensitivity even for low phase differences [20].

2.2.4 Polarized Light

Reflection microscopy with polarized light is used in metallography for the observation of strongly anisotropic materials, and is extensively used for ore microscopy because it is without doubt an important tool for the optical identification of phases. In contrast to transmission optical microscopy, one of the most important phenomena is not accomplished in a reflected beam, namely that the light is completely linearly polarized and vibrates in the same plane over the entire field of view. Deviations from this requirement occur if the incident angle differs from 90°, which is always the case in the margins of the observed area. Moreover, on placing the mirror in the vertical illuminator the ray undergoes a phase displacement differing from 0 to 180°, which also yields elliptically polarized light to a certain extent. These aberrations can be avoided by the use of the Berek compensation prism instead of a plane mirror in the vertical illuminator (see Fig. 9).

If the incident beam is unpolarized but the reflected beam is monitored with the inserted analyzer, its polarization state can be observed by rotating the stage and monitoring the reflectivity. Although the orientation-dependent variations in reflected intensity are only of the order of a few percent, isotropic and anisotropic materials can be distinguished. If rotated under plane-polarized light, anisotropic substances exhibit a variation in intensity and color of the reflected light (see Fig. 11). Their reflectivity varies between two extremes, which are called uniradial reflectivities. The difference between these extremes is referred to as bireflectance or double reflectivity. The bireflectivity approaches zero if the section of the crystal is normal to the optical axes or the winding axes or to another special cut. Reflectivity, bireflectivity, and their dependence on

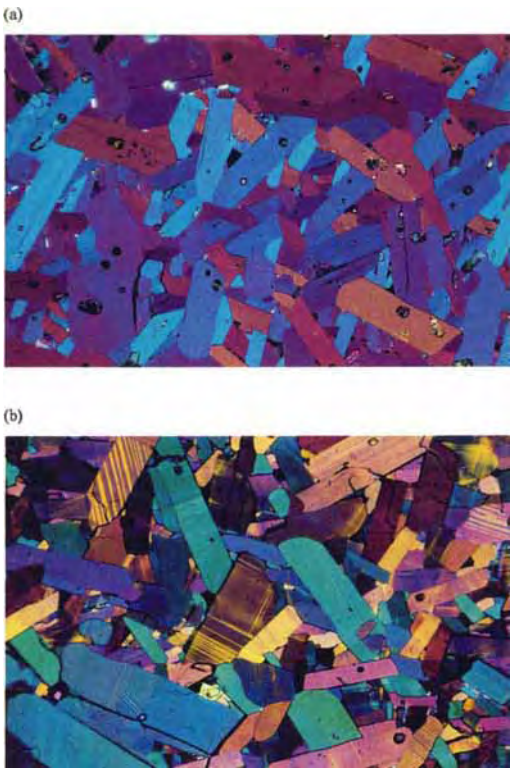


Figure 11. $YBa_2Cu_3O_{7-x}$ superconductor material, Fe/O_2 contrasted: (a) tetragonal phase and (b) orthorhombic phase with multiple twinning. $\lambda/4$ plate inserted; magnification, 100×. (Courtesy of U. Schäfer and H. Aslan, Max-Planck-Institute for Metals Research, Stuttgart, FRG.)

wavelengths are significant optical constants and can be measured by means of microphotometers. In the same way, the color may vary due to anisotropic absorption (see Sec. 1.2.4 of this Chapter). This effect is called reflection pleochroism, and can be a very typical characteristic of certain phases.

Under crossed polarizers, isotropic and anisotropic crystals can be distinguished in the same way as with only one polarizer, but the anisotropy effects appear much clearer. Isotropic material generally remains dark whereas anisotropic matter shows azimuth-dependent variations in reflectivity. Since reflection of absorbing matter usually yields slightly elliptically polarized light, even if the incident ray is completely plane polarized, rotating a crystal about 360° results in four positions of minimum reflectivity but not of complete extinction. As already stated above for bireflectance, specially oriented sections may also show quasi-isotropic behavior. Spectacular effects of reflection pleochroism may be obtained in some colored substances, which can even be intensified by using oil immersion.

If there is a magnetization perpendicular to the surface of the sample, the magnetic domains can be visualized. The contrast depends on the magnetization direction. Adjacent domains can be differentiated by applying a compensator with a variable phase shift.

2.2.5 Differential Interference Contrast

Differential interference contrast is one of the most widely applied techniques in reflection microscopy. It is more sensitive to local phase shifts than the phase contrast method and is therefore most commonly preferred. Moreover, its realization is rather simple and the images are well contrasted and in most cases easy to interpret. The principle can be adapted from that of transmission microscopy (see Sec. 2.1.5 of this Chapter) and consists of a double quartz prism, such as a Wollaston prism, between the vertical illuminator and the objective lens while employing cross-polarized light. This produces interference effects, to some extent similar to those obtained by oblique illumination. Color is produced by adding a sensitive tint plate. Adjustment of the prism produces changes in color. The optical elements for the reflection case are arranged as shown in Fig. 12.

Contrast is achieved for phase objects (see Sec. 2.1.3 of this Chapter) as in the transmission case, but with reference to the surface relief, which has to be considered for image interpretation (Fig. 13). Surface reliefs are commonly produced by chemical relief etching, by mechanical relief polishing as well as other physical procedures of sample preparation. However, surface reliefs also occur during growth and deformation of materials (growth steps, slip lines, etc.). The technique is also applied for the control of surface quality after polishing, grinding, etching, sputtering, evaporation, etc.

The image contrast cannot be directly converted into the dimension of a surface relief without a standard, as is possible in interferometry (see Sec. 2.2.6 of this Chapter). Nevertheless, inclinations of surfaces can be measured with respect to a standard plane. Because the optical system produces a less defined transfer of the phase shifts,

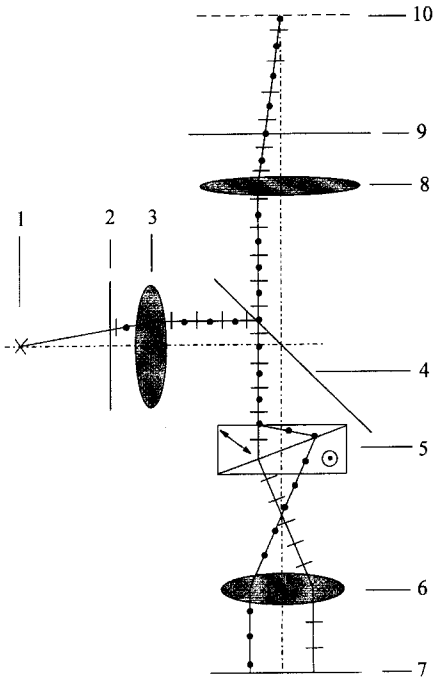


Figure 12. Arrangement of the optical elements for differential interference contrast according to Nomarski. In contrast to the transmission case (see Fig. 8) only one Wollaston prism is necessary, since beam splitting and beam junction take place in the same prism. 1, light source; 2, polarizer; 3, illumination lens; 4, semitransparent plane glass mirror; 5, Wollaston prism; 6, objective; 7, plane of object; 8, tube lens; 9, analyzer; 10, back focal plane.

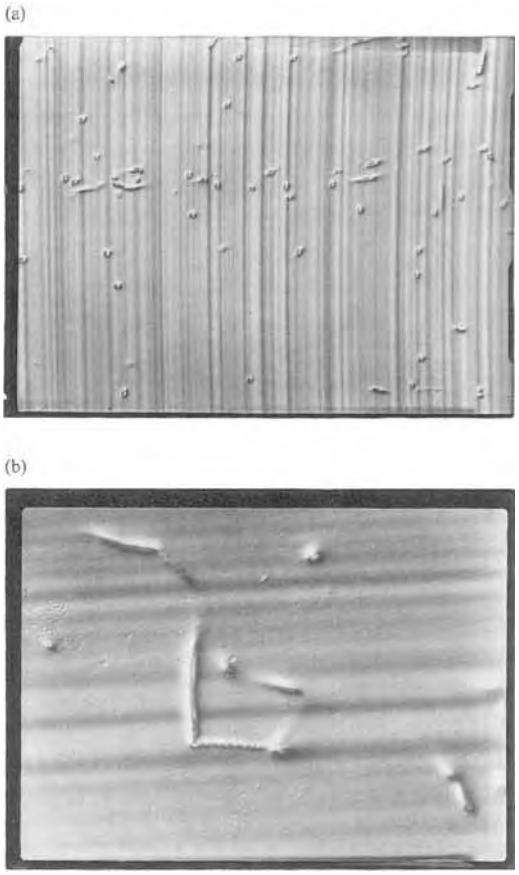


Figure 13. GaAs single crystal surface after photoetching (orientation {001}) visualized using differential interference contrast (Nomarski). (a) Overview (magnification, 50×), vertical striations are attributed to periodical fluctuations of the silicon dopant producing waviness during propagation. (b) Detail of (a) (magnification, 150×): line shape 'mountain chain' contrast of local dislocation paths and even pinning precipitations are seen (note small etch pits along a 'mountain chain'). (Courtesy of O. Oettel, TU Mining Academy Freiberg, FRG.)

no quantitative measurement of phase differences can be recommended. An excellent sensitivity of adjacent phase shifts as low as 1 nm (only in the direction of lateral beam splitting, as mentioned above) can be exploited with this technique.

2.2.6 Interferometers

The optical arrangement described in Sec. 2.2.5 of this Chapter can also be used for larger beam splittings with values above the resolution limit. In this case, a polarization interferometer is designed in which the surface can influence both beams. Such interferometers are stable and easy to handle. Quantitative investigations of phase shifts are carried out measuring the excursion of the interference fringes (Fig. 14).

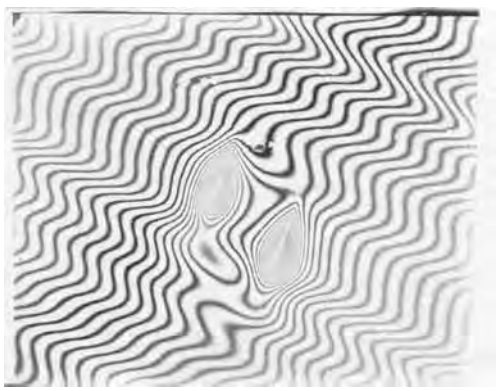


Figure 14. GaAs single crystal surface after photoetching (orientation {001}) visualized using total beam splitting (shearing mode) for the measurement of 'hillock' heights. Monochromatic light; magnification, 250×. (Courtesy of O. Oettel, TU Mining Academy Freiberg, FRG.)

However, higher precision can be achieved if the reference beam is not influenced by the object itself. A typical design is the Michelson interferometer (see Fig. 15).

Light from the monochromatic source (1) propagates in parallel after the collimator (2), and is split into two beams at the beam splitter (3). Both beams return to the beam splitter after reflection at the sample surface and at the reference surface, respectively. After a repeated splitting at 3, parts of the sample beam and of the reference beam which propagate toward the objective lens (6) can interfere because they are coherent. The interference figure therefore depends on the surface of the sample as well as on the inclination between the sample and the reference surface. Usually, a system of interference fringes is established where the distance between the fringes is controlled by the inclination between the sample and the reference surface. If no inclination exists, the image appears homogeneous and free of fringes. Deviations from evenness of the sample appear as local shifts of the fringes, which can be detected in order to measure the differences in height (see Figs 16 and 17).

Precise measurements can also be carried out by the compensation of the fringe shifts with additional compensation wedges or with parallel shift of the

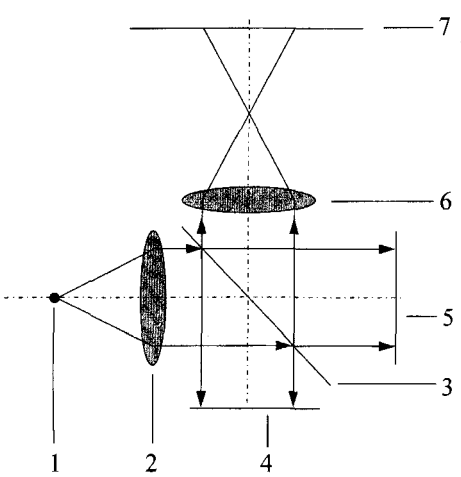


Figure 15. Principle of the Michelson interferometer. 1, monochromatic light source; 2, collimation, lens; 3, beam splitter; 4, sample surface; 5, reference surface; 6, objective; 7, image plane.

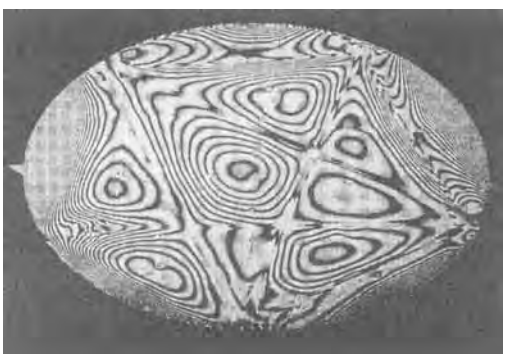


Figure 16. Interferogram of the thermal deformation of an antenna consisting of a carbon fiber-reinforced plastic, deformed by 2 K. (Courtesy of Steinbichler Optotechnik GmbH, Neubeuern, FRG.)

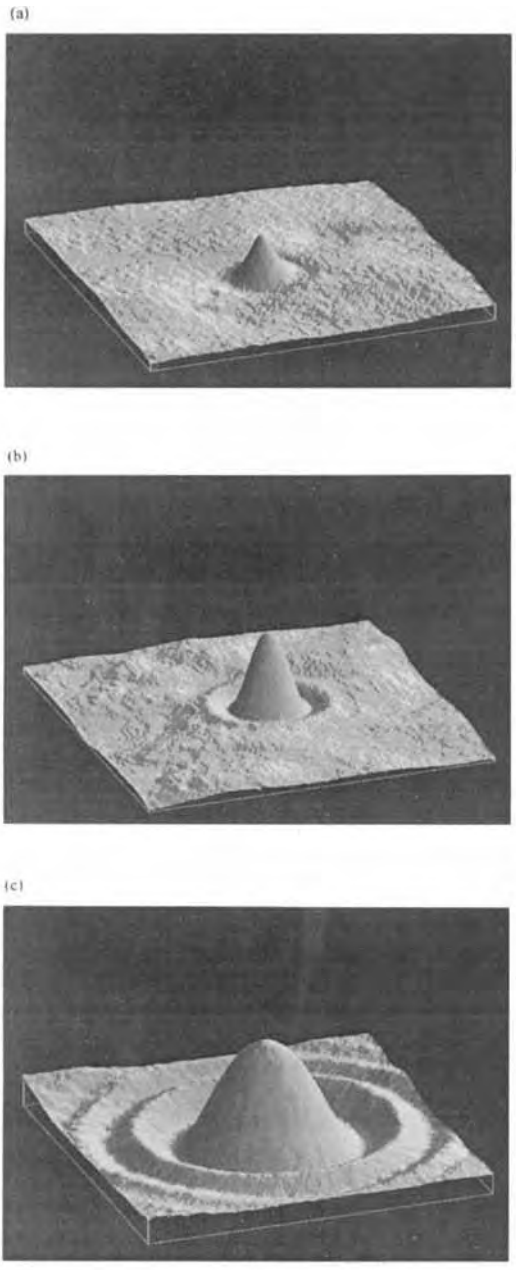


Figure 17. Shock propagation on a plate (45 × 45 cm). The three-dimensional images show a sequence of deformation hillocks by using interferometry. The hillock heights are (a) 0.7 μm, (b) 1.1 μm, and (c) 1.2 μm. (Courtesy of Steinbichler Optotechnik GmbH, Neubeuern, FRG.)

reference plane (see Fig. 15 (5)). Another variation of this technique which allows the application of higher objective aperatures was proposed by Linnik [21].

2.2.7 Interference Layer Contrasting

Anisotropic properties of materials with low relectivity can, however, be made visible by the technique of interference layer contrasting. Contrast-enhancing

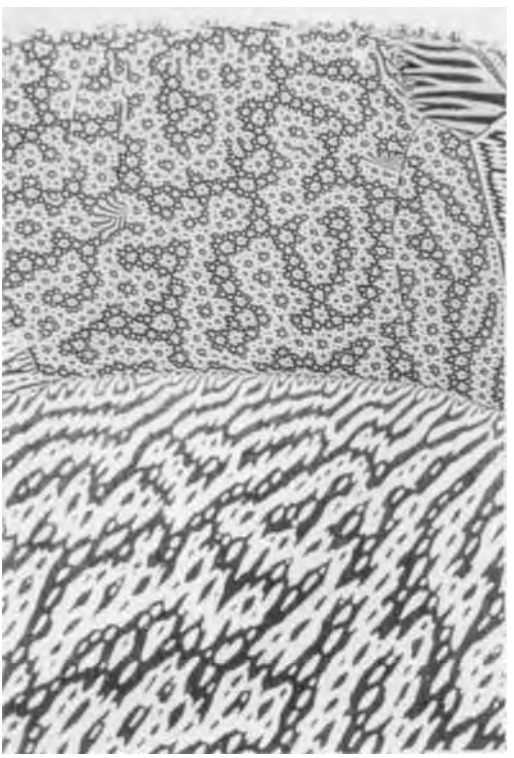


Figure 18. Microstructure of magenic domains in $SmCo_2$ (coarse domains) and $SmCo_{17}$ (fine domains) after sputtering with ZnSe. Polarized light; magnification, 100×. (Courtesy of U. Schäfer and G. Schneider, Max-Planck-Institute for Metals Research, Stuttgart, FRG.)

layers can be deposited on the surface of polished sections by reactive sputtering from a metallic cathode (Pt, Au, Pb, Fe, Al, etc.) in an oxygen atmosphere or by evaporation (ZnS, ZnSe, or ZnTe) (Fig. 18).

The possible methods for the deposition of coloring coatings are: (1) heat tinting, that is, exposing the sample to air at elevated temperatures in order to grow epitaxially thin oxide layers on certain preferential metallic phases which are sensitive to corrosion [22]; (2) color etching, which involves a reactive solvent-assisted deposition of precipitation layers; and (3) anodizing, meaning an electrolytic process for reactive film deposition during electropolishing (see Sec. 2.3 of this Chapter).

The optical effect of contrasting is caused by multiple relections of the incident beam at the sample/coating and coating/air interfaces, relations which introduce path differences as a function of layer thickness, refractive indices, and absorption coefficients of both the sample and the coating material. Certain wavelengths are then extinguished by interference, and the resulting complementary colors are observed. Since the polarization mode of the reflected beam, as well as the optical constants which are effective on the particular crystal surface, depend on the crystal orientation, strong contrasts may be observed even in single-phase materials. Because of their poor reflectivity and comparatively low refractive indices, the effects for ceramics are not as good as for metals, but satisfactory results can be obtained if the following conditions are observed: the ceramic phases should exhibit a high absorption, and opaque materials are preferred; platinum, iron, lead, or cryolithe (Na_3AlF_6) are the most suitable coating materials; and the thickness should not exceed 5 nm.

2.3 Remarks on Sample Preparation

To determine the structure of a natural or artificial material it has to be prepared for microscopic examination. The aim of proper preparation is to present the true structure of the sample. Sometimes the samples may be viewed without mechanical preparation. For transmission microscopy, the preparation of thin plane parallel sections is essential (see Sec. 2.1 of this Chapter), whereas in reflection microscopy only the sample surface is examined, and provides all the information from microstructural differences in reflection and absorption as well as phase shifts, which can be generated additionally during preparation by means of selective excavation (relief preparation).

Generally, all structural elements should be retained, the surface should be free of scratches and deformation, and no foreign matter should be introduced into the surface during preparation (Fig. 19).

The preparation route very much depends on the composition, the treatment and conditioning of the materials. For example, it is not sufficient to know that a sample is silicon nitride, but the preparer also has to know whether it is sintered or not, as well as whether it is expected to have porosity and impurities. These properties greatly influence the preparation and give an indication of the successful preparation route. An extensive description of efficient sample preparation routes for metals,

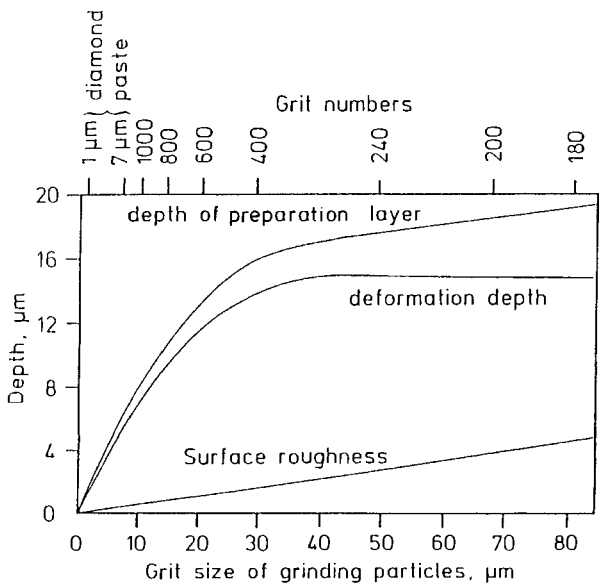


Figure 19. Reduction of surface roughness and deformation depth as a function of grit size of grinding materials for the example of steel. (Redrawn from G. Petzow, H. E. Exner in: *Handbuch der Mikroskopie in der Technik* (Ed.: H. Freund), Umschau-Verlag, Frankfurt **1968,** Vol. III, p. 1.)

alloys, glasses, ceramics, plastics as well as composites and coatings is summarized in [23]. Also included in this comprehensive handbook are detailed reference lists for the different groups of materials.

Acknowledgment

The valuable assistance of Mr. P. Dörnach in the preparation of the final manuscript is sincerely acknowledged.

2.4 References

[1] E. E. Wahlstrom, *Optical Crystallography*, 5th ed., Wiley, New York **1979**.
[2] A. Köhler, *Z. Wiss. Mikrosk.* **1893,** *10*, 433.
[3] A. Thaer in: *Handbuch der Mikroskopie in der Technik* (Ed.: H. Freund), Umschau-Verlag, Frankfurt **1968,** Vol. IV, p. 3.
[4] F. Jeglitsch in: *Handbuch der Mikroskopie in der Technik* (Ed.: H. Freund), Umschau-Verlag, Frankfurt **1968,** Vol. III, p. 1.
[5] F. Zernike, *Z. Phys.* **1935,** *36*, 848.
[6] F. Zernike, *Physica* **1942,** *9*, 686, 974.
[7] F. Zernike, *Phys. Blätter* **1955,** 159.
[8] C. W. Correns, *Fortschr. Min. Krist. Petrogr.* **1930,** *14*, 26.
[9] H. Wolter, *Fortschr. Chem. Forsch.* **1954,** *3*, 1.
[10] P. H. Ribbe, H. C. van Cott, *Can. Mineral.* **1962,** *7*, 278.
[11] H. Pichler, C. Schmitt-Riegraf, *Gesteinsbildende Mineralien im Dünnschliff*, Fedinand Enke, Stuttgart **1987**.
[12] D. J. Clinton, R. Morell, M. McNamee, *Br. Ceram. Trans. J.* **1986,** *85*, 175.
[13] H. Rosenbusch, E. A. Wülfling, *Microscopical Physiographic I*, Part 1, Nägele und Obermiller, Stuttgart **1921–4**.
[14] H. Winchell, *Am. Mineral.* **1946,** *31*, 43.
[15] A. C. Tobi, *Am. Mineral.* **1964,** *41*, 516.
[16] A. G. King, *J. Am. Ceram. Soc.* **1971,** *43*(7), 424.
[17] M. Francon, *Mikroskopie* **1953,** *8*, 260.
[18] G. Nomarski, French Patent **1952,** *1059*, 123.
[19] F. H. Smith, *Research* **1955,** *8*, 385.
[20] H. Beyer in: *Handbuch der Mikroskopie*, 3rd ed. (Ed.: H. Riesenberg), Technik, Berlin **1988,** Chap. 6.
[21] W. Linnik, *C. R. Acad. Sci.* **1933,** *1*, 18.
[22] J. V. Cathcart, G. F. Peterson, C. J. Sparks in: *Surface and Interior Chemical and Physical Characteristics* (Eds.: C. J. Burke, T. A. Reed, G. A. Weiss), Syracuse University Press, Syracuse **1967**.
[23] G. Petzow, *Metallographisches, Keramographisches und Plastographisches Ätzen*, Borntträger, Berlin **1994**.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

3 Raman Microscopy

3.1 Introduction

From its discovery in 1928 until about 1975, Raman scattering has been used for studying bulk samples of macroscopic dimension, furnishing information about fundamental molecular properties and providing an important part of laboratory spectroscopy. Raman spectroscopy is based on the spectral distribution of inelastically scattered light and is a highly selective technique for investigating molecular species in all phases of matter, as they are fingerprinted by their vibrational spectra. The introduction of laser sources has greatly enhanced the utility of Raman spectroscopy for microanalysis. Indeed, Raman scattering can provide information which previously was not available from any other widely used techniques such as electron, ion, and Auger microprobes. These latter techniques can readily identify, map out the distribution, and determine the quantity of elemental constituents present but they do not directly distinguish the chemical forms of polyatomic species present as specific compounds in a sample. Using lasers as excitation sources for Raman scattering, along with the ongoing development of instrumentation for optical spectroscopy, means that Raman microspectrometry has now matured to the point at which nondestructive chemical microanalysis and mapping has become routine practice for both research and industrial purposes. This paper describes recent developments of confocal Raman microspectrometry and imaging.

3.2 Principles of Raman Spectroscopy

Measurements performed with Raman spectrometers are mainly based on the excitation and detection of the normal (spontaneous) Raman effect, which has been fully described [1]. The effect is an inelastic scattering process, observed in the near UV-vis to the near-IR spectral region, which involves the interaction of a monochromatic beam of light with the molecules of the sample. This interaction produces scattered radiation at different frequencies. These frequency shifts (from the exciting line) are identified with the frequencies of the atom oscillations in polyatomic structures contained in the sample. A schematic representation of a Raman spectrum is presented in Fig. 1. Excitation with the green line (514.5 nm) of an argon ion laser is assumed. The Stokes Raman lines

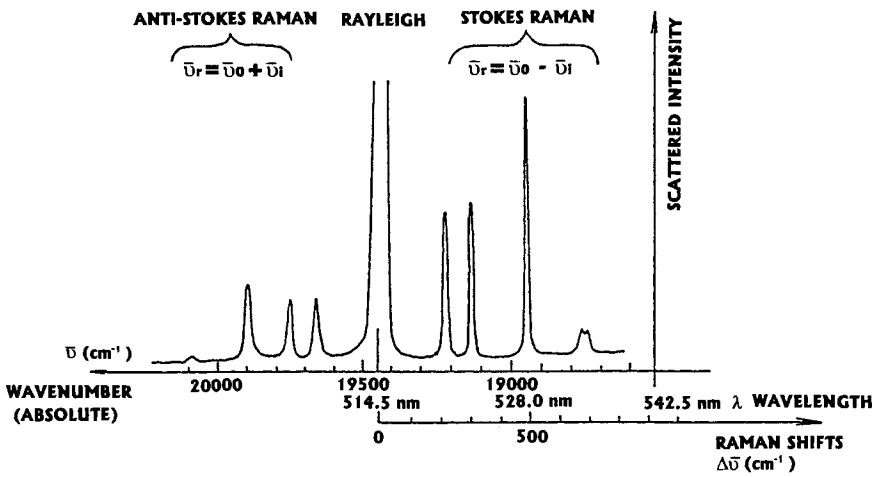


Figure 1. Schematic representation of a Raman spectrum excited with the green line of an argon ion laser, $\lambda_0 = 514.5\,\text{nm}$.

$(\bar{\nu}_0 - \bar{\nu}_i)$ and the corresponding weaker anti-Stokes lines $(\bar{\nu}_0 + \bar{\nu}_i)$ appear symmetrically on both sides of the strong Rayleigh line $(\bar{\nu}_0)$. In this representation, $\bar{\nu}_0$ and $\bar{\nu}_i$ are the absolute wave numbers $(\bar{\nu} = \nu/c)$ expressed in cm^{-1} of the exciting line and of the Raman lines, respectively. The Raman shifts $\bar{\nu}_r$ are read directly on the Raman spectrum recording, which usually consists only of the more intense Stokes part. The intensities of the lines are determined by the Raman cross-section (see below) and are directly proportional to the number of molecules which are probed. The kind of information provided by the Raman spectrum is essentially the same as that obtained from infrared spectra. Thus, the Raman spectrum can be regarded as a unique fingerprint which also contains information on the local molecular environment (e.g., amorphous or crystalline phases). In the spontaneous Raman effect (described above) the incident photon energy is below the energies of any excited electronic levels. However, if the exciting wavelength is such that the incident photon energy approaches or is equivalent to these levels, then other inelastic processes, such as resonant Raman scattering [2] and fluorescence [3], may be induced. Fluorescence emission, which is often an order of magnitude or more than the Raman scattering, may sometimes mask the Raman emission. This is one reason why Raman spectroscopy is not commonly used in industry where most samples exhibit a strong fluorescence background. Nevertheless, fluorescence emission can be minimized by using near infrared excitation (i.e., $\lambda > 1000\,\text{nm}$).

3.3 Confocal Laser Raman Microspectrometry

3.3.1 Instrumentation

The intensity of the signal delivered by the detector of a spectrometer analyzing a

given Raman line at the wavelength λ can be expressed by

$$S \sim I_0 \sigma_\lambda N \Omega T_\lambda s_\lambda \qquad (1)$$

where I_0 is the laser irradiance at the sample ($W\,cm^{-2}$), σ_λ is the differential cross-section for the Raman line analyzed (cm^2 $sterad^{-1}$ $molecule^{-1}$), N is the number of molecules in the probed volume V, Ω is the solid angle of collection of the Raman light, and T_λ and s_λ are the throughput of the instrument and the sensitivity of the detector at λ, respectively. When a small volume of matter has to be examined, only a few parameters can be modified to compensate for the large reduction in the number of molecules, N, in the probed volume, V, namely I_0 and Ω. In fact, the probed volume and the solid angle of collection are not independent and extensive work at the Laboratoire de Spectrochimie Infrarouge et Raman (LASIR), several years ago, was devoted to exploring techniques for the development of micro-Raman instruments [4, 5]. From this experience it was concluded that the use of microscope objectives for both illuminating the sample and collecting the Raman light was the best way to decrease V while increasing Ω and I_0. Indeed, microscope objectives, which are high-numerical aperture (N_A) optics, are able to focus the laser beam into a very small volume and to collect, under a wide angle, the light scattered by this volume. Thus, both the enormous increase of the local irradiance I_0 and the wide angle of collection Ω compensate for the decrease of the number of molecules N. A microRaman instrument possesses a good quality light microscope coupled to spectrographs (or spectrometers). This optical coupling must be optimized from the sample to the photoelectric detector via the collection optics and the spectrograph and the spectrometer. A schematic diagram of the widely adopted laser focusing, sample viewing, and scattered light collection geometry is represented in Fig. 2. It can be seen, from this figure, that a microRaman

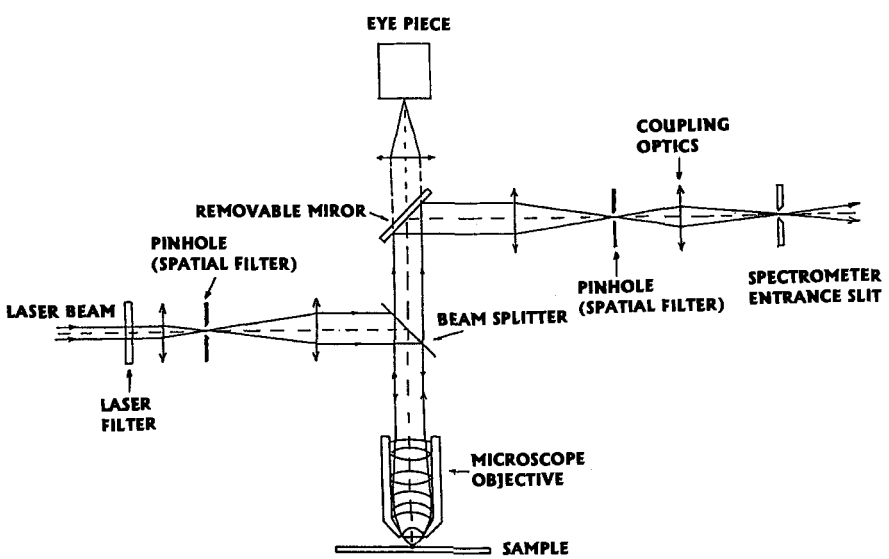


Figure 2. Optical scheme of the illuminating and collecting optics of a micro-Raman spectrometer.

spectrometer was conceived, at its origin, as a confocal microscope. In fact, a real confocal configuration requires a very accurate optical alignment and a high degree of stability and reproducibility of the mechanical and optical alignment.

3.3.2 Confocal Raman Microscopy

The major advance in Raman microprobing during the last few years deals with the extension of the principles of confocal microscopy. Applied to Raman microprobing, the confocal technique provides an efficient way to obtain interference-free Raman spectra as well as 2D or 3D selective images of small specimens embedded within strong scatterers or fluorescent transparent media. Some account of the advantages of confocal microscopy and its benefit to Raman microprobing are now presented.

Optical scanning microscopes, using a laser point source sharply focused in a diffraction limited spot on the specimen, have been developed to significantly improve the spatial resolution [6]. This principle, first applied to transmitted or reflected light measurements, has been further extended to microfluorimetry [7].

Unlike the conventional microscope, where the entire field of view is uniformly illuminated and observed, the 'confocal' arrangement takes benefit from the combined effects of spatial filtering by optically conjugated pinhole diaphragms. Such a system isolates the light originating from a small region of the sample coincident with the illuminated spot, and efficiently eliminates the contributions from out-of-focus zones.

The main advantages of confocal microscopy may be summarized as follows:

(i) There is a slight improvement of lateral resolution. The intensity point spread function (PSF) for the conventional microscope is given by

$$|h_1(v)|^2 = [2J_1(v)/v]^2 \tag{2}$$

whereas for the confocal microscope it becomes

$$|h_1(v)|^2 = [2J_1(v)/v]^4 \tag{3}$$

where J_1 and v are the first order Bessel function and a normalized distance to the optic axis, respectively.

The PSF, otherwise limited to the Airy diffraction disc dimension, can be reduced theoretically by a factor 0.7 but at the expense of very low throughput, since it requires pinhole diaphragms of small diameter.

(ii) There is a dramatic improvement of the depth discrimination which allows an 'optical sectioning' of transparent specimens. A complete calculation of the depth of focus can be found in the literature [6, 8–10]. A careful evaluation of the spatial distribution of local irradiance in the focal volume close to the laser beam secondary waist, combined with the calculation of light collection by a wide aperture objective, are necessary to evaluate correctly the depth discrimination. The intensity variation along the z axis in the neighborhood of the image plane of a point object placed at a nomalized distance $u = (2\pi/\lambda)z\sin^2\alpha$ from the focal plane of the objective (of angular aperture α) is given by

$$I_1(u) = |h_1(u)|^2 = |(\sin u/4)/(u/4)|^2 \tag{4}$$

In a confocal microscope the intensity variation depends on the properties of

both the condenser and the objective lens. The intensity profiles vary according to the relationship:

$$I_1(u) = |h_1(u)h_2(u)|^2 = |(\sin u/4)/(u/4)|^4 \quad (5)$$

To remember the physical limit imposed in confocal microscopy, an approximate expression of the depth of focus Δz has been proposed:

$$\Delta z \geqslant \pm \frac{4.4n\lambda}{2\pi(NA)^2} \quad (6)$$

where NA is the numerical aperture of the objective. For example, a 'dry' objective with $NA = 0.92$ would produce a depth discrimination $\Delta z = 0.6\,\mu m$ at $\lambda = 514.5\,nm$. (Such a physical limit is evidently not compatible with the 100 nm which may be found in some optimistic advertisement.)

It is worth noting that the throughput of the confocal systems proposed for transmitted or reflected light microscopy is very low compared to conventional techniques. The overall transmission of the whole optical system is typically of the order of 1 to 5%. The extension of the principles of confocal microscopy to Raman microprobing is not straightforward and necessitates a careful investigation of the specific requirements of this technique.

(iii) There is an extremely low light level of the Raman signals compared to the intensity of the excitation radiation which is retroscattered by reflection at the interfaces or that of the diffuse scattering by the specimen (typically 10^{-3} to 10^{-12}). Consequently, particular attention must be paid to the rejection of stray light in the whole optical system, including the spectral analyzer, together with a high throughput and high performance photodetectors.

(iv) The possibility of spectral analysis in a relatively narrow domain adjacent to the intense parent line at the laser frequency (typically 100 to $3500\,cm^{-1}$ for routine vibrational spectroscopy).

(v) The spectral resolution is much higher than for absorption or fluorescence (typically 1 to $10\,cm^{-1}$).

(vi) The best conditions of measurement are formed by a proper choice of the laser wavelength which may be varied from near UV to near IR according to the necessary balance between resonance Raman enhancement and superposition of luminescence bands, which would otherwise severely impair the detection of weak Raman lines.

As a result of a close collaboration with our laboratory, a versatile laser Raman confocal system has been developed by Dilor (Lille, France) [11] in which the following critical parameters have been optimized:

- laser beam spectral and spatial filtering
- beam matching of both the laser source and the spectrometer, to the pupil apertures of various objectives (macro and micro) by means of afocal expanders
- exact optical conjugation of the illumination and measurement spatial filters with the observed zone of the sample
- parallel light beamsplitter, used in a proprietary configuration to couple the system with two macro or micro entrances without optical switching
- easy control of all adjustments and precise selection of the observed area by several CCD microcameras

- variable aperture of the pinhole diaphragm, which directly determines the spatial resolution.

As expected, Raman measurements take full benefit of the dramatic improvement of spatial resolution, without significant loss of detectivity, provided that the accuracy, stability and reproducibility of all adjustments are effectively realized.

A calculation of the proportion of the flux, Φ_T, which is transmitted through the pinhole diaphragm (Raman intensity), has been made. This is given by

$$\Phi_T \approx \frac{R_d}{\{\Gamma_0[NA\,z + R_0\sqrt{1+(\sin^{-1}NA\,z)^2}]\}^2} \tag{7}$$

where z, R_d, R_0, Γ_0 and NA are the axial altitude of a slice of the sample, the radius of the pinhole, the radius of the laser beam at sample, the magnification factor and the numerical aperture of the microscope objective, respectively. As a result, it is worth noting that three main parameters are accessible to the analyst for the control of spatial resolution: namely, the numerical aperture of the objective, the total magnification of the optical system, and the dimension of the pinhole diaphragms.

The 'optical sectioning' capability is concerned with the shape of the depth discrimination function as given in Fig. 3. In practice, this function is usually measured by the full width at half maximum (FWHM) of the plot of the Raman intensity of a thin layer specimen of thickness dz versus its axial displacement z.

The improvement of axial resolution certainly offers the most attractive advantages for microRaman analytical applications. Consequently the ability of the confocal system to virtually eliminate the out-of-focus regions of the sample, whose contribution to the signals are negligible, has decisive advantages, including: contrast enhancement, discrimination of a well defined spatial region in a

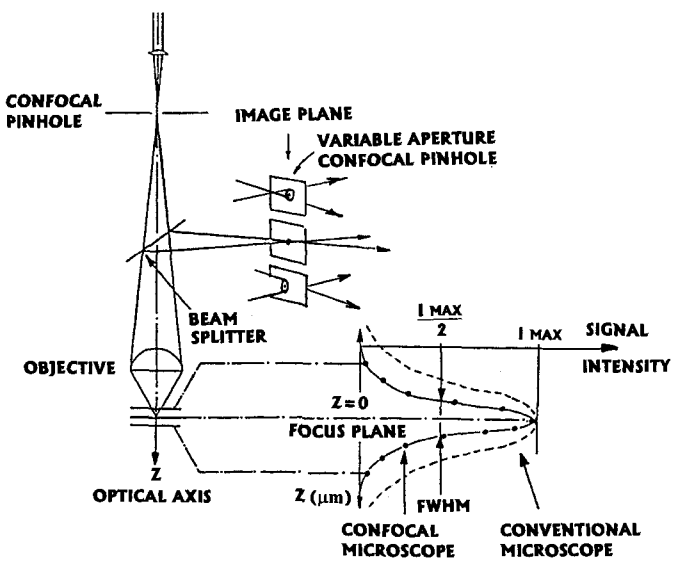


Figure 3. Principle of confocal microscopy and depth discrimination.

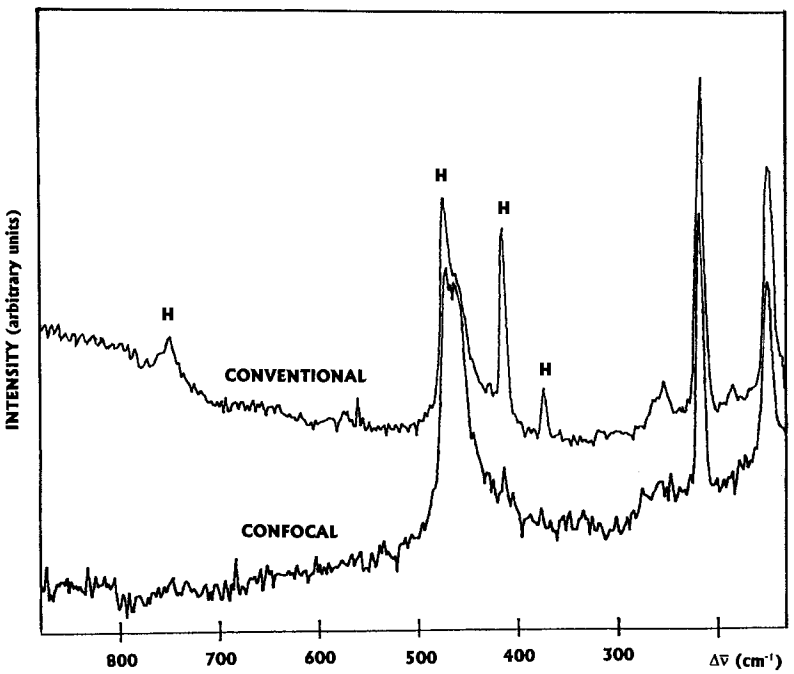


Figure 4. Illustration of the depth discrimination capability. Solid inclusion of chalcopyrite within a ruby host.

complex multiphase specimen, rejection of stray light, and rejection of the surrounding material when small inclusions embedded in a matrix or immersed in a liquid are observed. An illustration of the confocal axial discrimination is given in Fig. 4.

The spatial filters are also optically conjugated to the entrance slit of the spectrometer and coupled to the array of pixels of the multichannel photodetector (CCD or diode array) so that the entire system has to be conceived and optimized to ensure the best performance. This gives a clear superiority of dispersive spectral analyzers over Fourier transform interferometers in this matter because the point illumination of the narrow slit of a stigmatic spectrograph is well adapted to the spatial filter whereas the Jacquinot advantage of a Michelson interferometer is lost when the entrance aperture is reduced to a pinhole.

3.4 Raman Imaging

3.4.1 Basic Principles

As a result of major progress in the performances of low light-level detection systems, monochromators, and laser excitation schemes, the vibrational Raman spectra not only serve to identify, at the microscopic level, the various molecular compounds present within a heterogeneous microsample but also can furnish information about their spatial distribution with a resolution close to the diffraction limit.

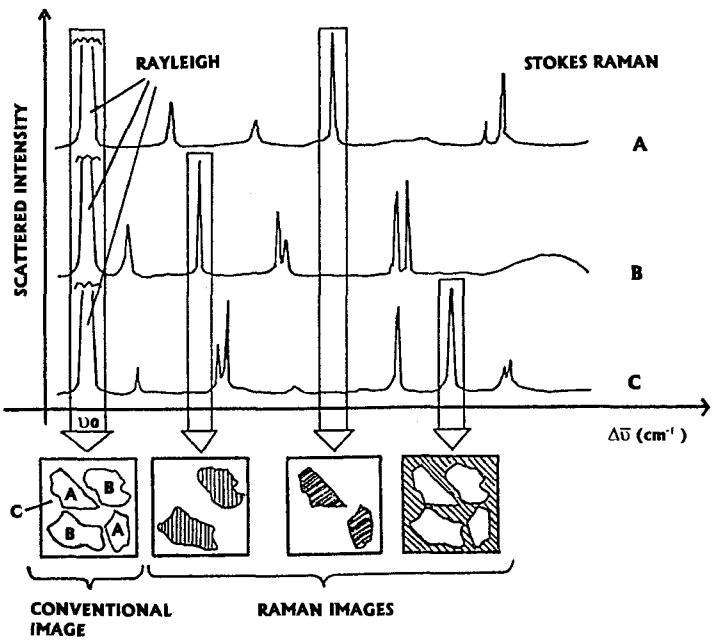


Figure 5. Principle of Raman imaging.

A selective image of the sample, depicting the distribution of a given molecular species, is obtained by passing the scattered radiation through a bandpass filter which isolates a narrow spectral region centered on a characteristic Raman frequency of this species (Fig. 5). By a judicious selection of nonoverlapping Raman lines, the spatial distribution of all the molecular compounds present in the specimen can be mapped out separately.

The production of microscopic 'molecular' maps using Raman signals can be realized by direct image forming procedures or alternatively by scanning techniques that are already described in the literature. In this description we simply wish to point out the advantages and major limitations of both procedures that led us to design a new concept of Raman imaging by benefitting from the advantages of confocal microscopy.

3.4.2 Various Mapping Configurations

3.4.2.1 Wide-Field Laser Illumination and the Direct Image Forming Procedure

Global laser illumination of a large area (typically 150–400 μm in diameter) of the sample within the field of view of an optical microscope was pioneered in the early history of Raman microscopy (Raman microprobe MOLE [12, 13]). As presented in Fig. 6, by a proper coupling of the illuminated area to a 2D multichannel detector (TV camera) through a concave grating double monochromator acting as a bandpass filter, the potential for direct imaging of the spatial distribution of molecular species in the sample has been demonstrated. In this configuration both dimensions of the detector collect spatial

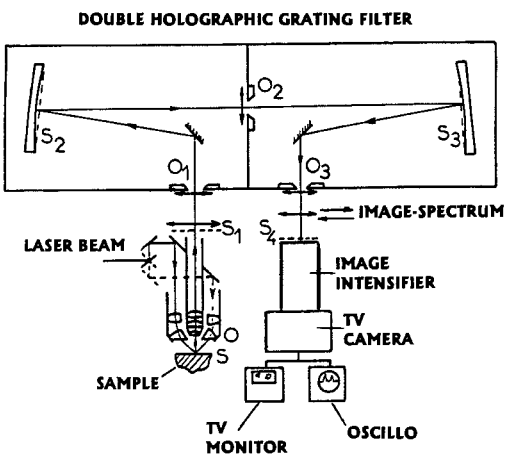


Figure 6. Optical scheme of the MOLE instrument in the Raman imaging mode.

information and only one wavelength at a time can be selected. However, to maintain the resolving power of the microscope objective along the whole optical system, the slit width of the monochromator must be as wide as the projection of the exit pupil of the microscope objective (typically a few mm in diameter). Therefore, the MOLE microprobe suffered from poor stray light rejection and low resolution performance (typically a few tens of cm^{-1}) as well as a lack of sensitivity that limited its application to a few number of samples consisting of fluorescence-free strong Raman scatterers.

Improved throughput and stray light rejection of global illumination microprobes are achieved by using a holographic notch filter for the rejection of the laser radiation in combination with a single grating spectrograph equipped with a CCD detector.

Replacing the grating spectrograph by a set of interference filters or a tunable acousto-optic filter to isolate a characteristic Raman band can result in a more compact design. Such instruments [14] with high optical throughput enable the relatively fast recording of widefield Raman or fluorescence images with good signal-to-noise ratio although they still suffer from the lack of depth resolution inherent in global illumination microscopy where confocality is absent. Some widefield illumination imaging systems are also based on Hadamard transform techniques for the recovery of both spatial and spectral data [15].

3.4.2.2 Digital Image Restoration Techniques

As we have seen previously, three-dimensional micro-Raman imaging is not readily achievable by direct imaging techniques due to the poor axial resolution of the conventional microscope which does not allow optical sectioning. With the development of powerful personal computers, an alternative to confocal microscopy can be the use of *a posteriori* data processing to improve the spatial resolution.

Significant improvement in the axial resolution of the conventional microscope has been obtained by sophisticated posterior digital image restoration processing. For instance, a nearest-neighbor deblurring technique to reject out-of-focus information and sharpen Raman images acquired by a Hadamard transform Raman microscope has been reported [16].

More recently a constrained iterative image restoration technique has been used with success to generate a stack of well-resolved Raman images which describe the three-dimensional topology of a sample [17].

3.4.2.3 Line Laser Illumination

Line laser illumination offers a further approach to Raman imaging together with the use of two-dimensional CCD detectors [18–21]. Unlike the direct image forming procedure, line illumination implies raster scanning of the laser beam onto the sample and computer-based reconstruction of 2D selective images from intensity profiles at several characteristic frequencies.

Conventional Line Scanning Imaging

Before entering the microscope objective the laser beam is scanned in the image plane of the microscope by a scanning device (either a mirror or a lens) driven by a signal generator delivering a triangular wave form. The laser beam is then focused by the microscope objective on the sample to scan a line whose length is adjustable. The scattered light along this line is then collected by the same microscope objective and brought to focus at the entrance slit of the spectrograph. A two-dimensional CCD detector at the exit port of the spectrograph simultaneously collects both spatial and spectral information originating from the illuminated line area of the sample (Fig. 7). Spatial information is contained in the X direction of the CCD parallel to the entrance slit while the perpendicular Y direction relates with spectral data. Two-dimensional images are acquired sequentially by moving the microscope stage step-by-step in the Y direction while storing, for each position of the sample, a complete set of intensity profiles at several wavenumbers. The omputer can then build up monochromatic images of the sample at any wavenumber by retrieving from its memory the

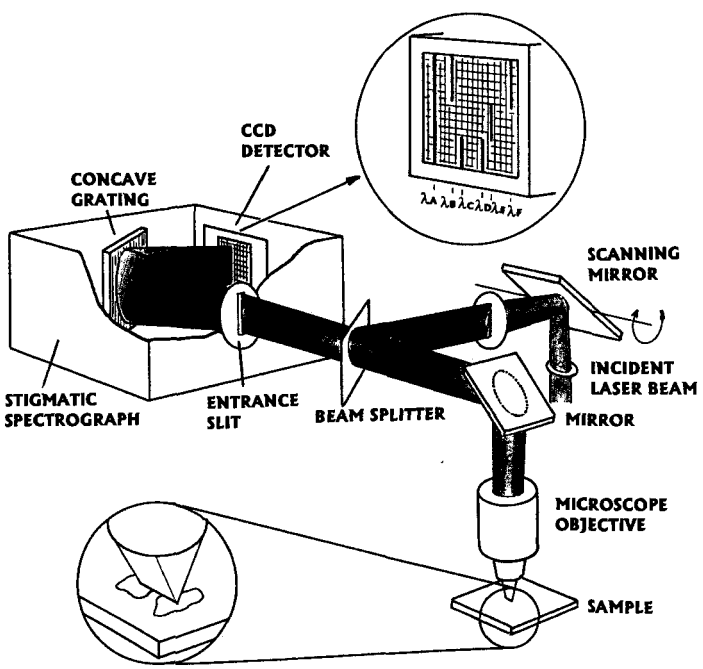


Figure 7. Laser Raman line scanning illumination optical scheme.

proper spectroscopic information for each location.

Compared with the direct imaging procedure, line scanning presents some advantages: namely, (i) an improved stray light rejection and an improved spectral resolution since the entrance slit is now optically conjugated to the narrow laser line, and (ii) a higher contrast due to a partial confocal effect at the entrance slit which acts as a one-dimensional spatial filter. However, it cannot fully benefit from the confocal advantage which requires a perfect spatial filtering not achievable by the scanned line configuration described above.

Confocal Raman Line Scanning Imaging

To take full advantage of the basic confocal principle we have developed the confocal Raman line scanning (CORALIS) optical configuration [22, 23] which is an improved design of the conventional line scanning technique. This optical arrangement (Fig. 8) allows line illumination of the sample and at the same time spatial filtering of the scattered signal for any

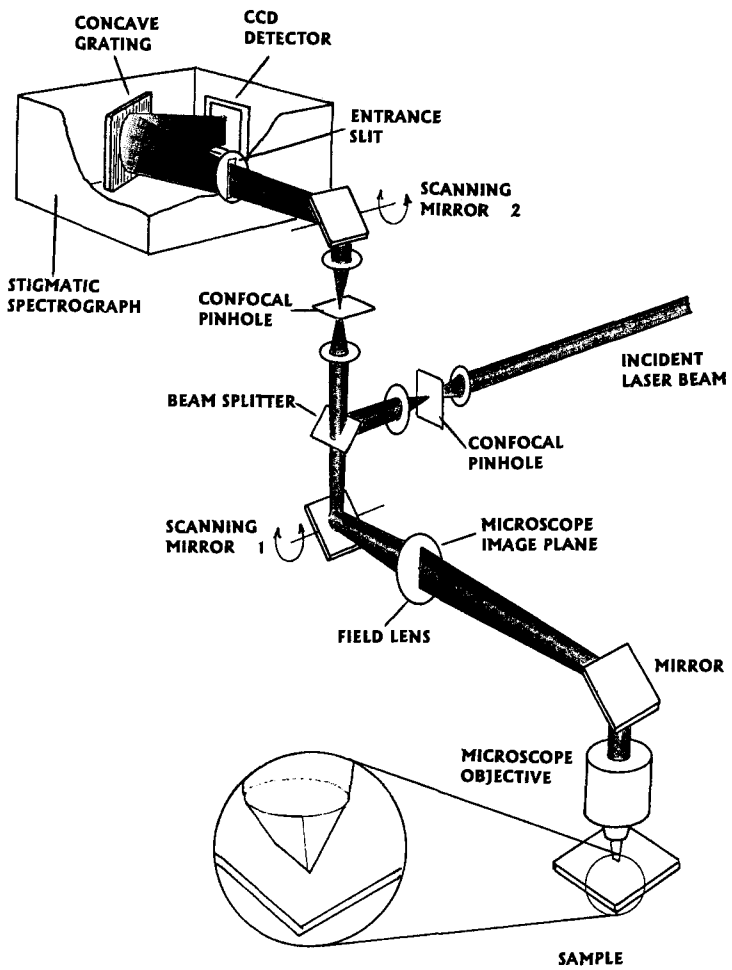


Figure 8. Confocal Raman laser line scanning (CORALIS) illumination optical scheme.

position of the laser beam onto the sample (confocal arrangement) by means of an adjustable confocal pinhole diaphragm. This is carried out by means of a two-fold scanning process.

The incident laser beam spatially filtered by a fixed aperture pinhole diaphragm (optically conjugated to the focal plane of the microscope) is focused in a narrow line in the image plane of the microscope by a first scanning mirror optically conjugated with the exit pupil of the microscope objective. For this purpose, the scanning mirror is imaged onto the exit aperture of the objective so that the laser beam always illuminates the full pupil independently of the scan angle of the scanning mirror. The microscope objective brings to focus the laser beam onto the sample located in the focal plane of the microscope. Since the scanning frequency is in the range of a few tens of oscillations per second (hertz) the sample seems continuously illuminated along a narrow line. However, at each scan angle of the scanning mirror corresponds a small circular region of the sample illuminated by the laser beam from which a scattered radiation is collected by the same objective. Since the path of light rays is reversible, for each position of the scanning mirror the scattered radiation exactly retraces the whole optical path back to the scanning mirror and the pinhole aperture. Actually, a beamsplitter located between the first scanning mirror and the pinhole aperture is used to separate the incident laser beam from the scattered beam. The laser beam is reflected by the beamsplitter while the scattered beam is sent toward the spectrograph through the beamsplitter. As the scanning mirror is used twice (on the incident beam and on the beam back to the sample) the scattered beam leaving the scanning mirror toward the spectrograph does not move anymore and can be focused on a second confocal pinhole, which is placed on the image of the first one in the reflection on the beamsplitter. Therefore, the signal from each illuminated point of the sample is filtered out by this adjustable pinhole and this configuration is strictly confocal, independently of the position of the scanning mirror.

After passing through the confocal pinhole the scattered light from a given sample location is deflected by a second scanning mirror (imperatively in phase with the first one but not necessarily with the same deviation) and focused on a given position of the entrance slit of the spectrograph. As the laser beam is scanned on the sample the scattered beam from all the points of the illuminated region of the sample is scanned over the slit so that one point of the sample is conjugated with a given point of the slit.

At the exit port of the stigmatic spectrograph a CCD detector accumulates the signal from several scans and then delivers a set of intensity profiles at various frequencies. Each profile depicts the spatial distribution of a given molecular species along the scanner laser line. Two-dimensional chemical images of the sample are acquired line-by-line by moving the sample step-by-step and storing all the data for each line. Reconstruction of three-dimensional images requires a further stepwise displacement of the sample along the optical axis. Powerful dedicated softwares have been especially developed to automate the successive phases of data collection and image reconstruction.

Besides the benefit of the confocal arrangement, which leads to a resolution

improvement and better depth discrimination, the CORALIS configuration provides an additional integrated zoom capability. As the two scanning mirrors can be operated independently from each other (although they must remain imperatively synchronous) a given area of the sample may be enlarged on the detector by only changing the scan angle of the 'laser' scanning mirror. Typically the scan angle of the slit scanning mirror (no. 2, Fig. 8) has not to be changed after it has been adjusted so that the scattered scan line just covers the spatial dimension of the CCD detector. On the contrary, for each sample, the laser scanning mirror (no. 1) has to be adjusted to illuminate only the region of interest that needs to be imaged. Actually, with a triangular command signal the line ends exhibit a higher intensity due to the change of the scanning direction. For this reason the laser scanned line is defined to cover a longer path than required while a mechanical mask eliminates the unwanted 'hot spots'. The zoom facility does not improve the overall resolution of the instrument which is still defined by the confocal configuration. However, it always ensures the best adaptation of the size of the sample to the size of the detector and can be used to increase the definition of the final image as characterized by the spread factor (number of pixel of the detector per micrometer of the sample).

With its two independently operated scanning elements the CORALIS configuration allows different modes of operation:

(i) Basic confocal microspectroscopy. This provides spectra of small samples with both scanning mirrors at rest.

(ii) Spectrum acquisition of fragile homogeneous samples. By scanning the laser beam onto the sample the incident laser power is spread over a large area so that sample degradation is reduced. By maintaining the slit scanning mirror at rest, the scattered signal from the whole illuminated region of the sample can be focused to a single point at the entrance slit in order to increase the intensity of the resulting spectrum.

(iii) Sample evolution versus physical parameter variation. During the integration period of the CCD detector, the scattered signal from a single point of the sample can be imaged on different parts of the entrance slit. Therefore, after the readout time, the detector delivers a set of spectra retracing the evolution of this point of the sample related to the change of any physical parameter (temperature, pressure, etc.).

(iv) Line profiling.

(v) 2D imaging.

(vi) Optical sectioning and 3D imaging Raman tomography.

Figure 9 illustrates the power of the CORALIS technique for Raman image reconstruction. This figure corresponds to the analysis of a part of the surface of a metal substrate on which a graphite layer has been deposited by chemical vapor deposition (CVD). It was possible to obtain two distinct Raman images by reconstructing these last images from the characteristic band of graphite despite the small downshift of the band due to a compressive stress induced during the deposit. In the image reconstruction process, the software extracts, from the collection of spectra corresponding to all the pixels (in this case, 625), the particular

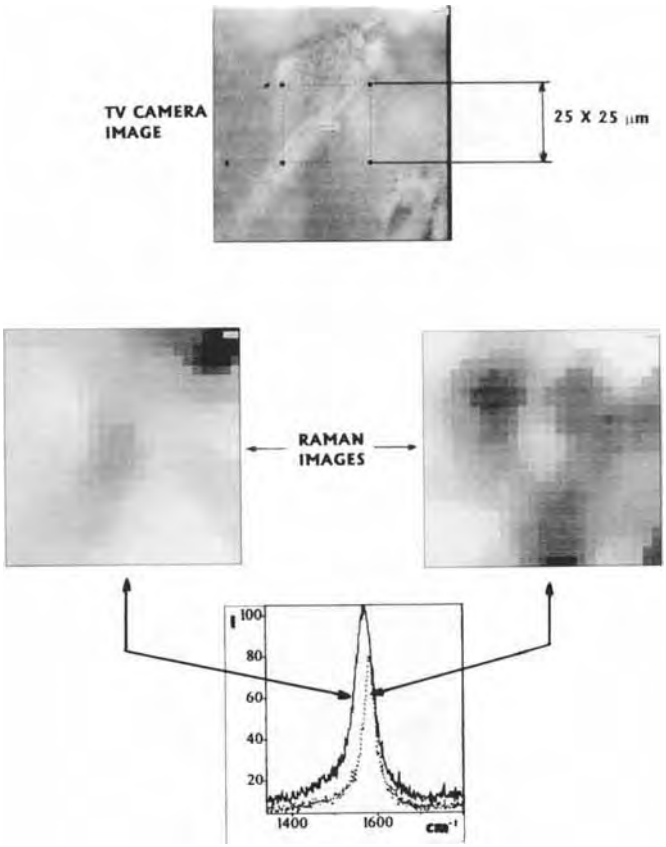


Figure 9. Illustration of the capability of the CORALIS technique to reconstruct Raman images from close Raman bands. Evidence of stress heterogeneities in a graphite layer deposited on a metal surface. Raman image resolution: 1 µm per pixel. Courtesy of DILOR (Lille, France) Application laboratory.

spectra for which the Raman bands about $1600\,cm^{-1}$ exactly fit the profile of each of the two shifted bands. It is impossible to produce such maps as evidence by using direct Raman imaging techniques, for which the difference between the wavenumbers of two close Raman bands has, of necessity, to be at least several tens of wavenumbers to give distinct images.

3.4.3 Future Developments

As with Raman microspectrometry, Raman imaging techniques suffer from fluorescence emission of most samples when the excitation wavelength is chosen in the visible region. A consequence of this is that most samples, especially those originating from industrial processes, cannot be studied because of their high level of fluorescence background. A significant progress in this field is expected when 2D near-infrared solid state detectors become commercially available since near infrared excitation drastically reduces the amount of fluorescence emission [24].

3.5 Conclusion

With routine detection limits in the nanogram range, high molecular selectivity and

high spatial resolution, microRaman spectrometry has now become a major microanalytical technique with research and industrial applications. Furthermore, the use of bidimensional CCD detectors have enhanced the capability of direct Raman imaging and confocal Raman imaging techniques to produce chemical maps.

3.6 References

[1] D. A. Long, *Raman Spectroscopy*, McGraw-Hill, New York **1977**.

[2] S. A. Asher, *Anal. Chem.* **1977,** *65*, 59A.

[3] M. Salmon, P. Viallet, *Innov. Tech. Biol. Med.* **1981,** *2*, 35.

[4] M. Delhaye, P. Dhamelincourt, *J. Raman Spectrosc.* **1975,** *3*, 33.

[5] P. Dhamelincourt, Doctoral Thesis **1979,** Université de Lille I, France.

[6] T. Wilson, C. Sheppard, *Theory and Practice of Scanning Optical Microscopy*, Academic Press, London **1984**.

[7] S. Kimura, C. Munakata, *Appl. Opt.* **1990,** *29*, 489.

[8] T. Wilson, *Confocal Microscopy*, Academic Press, London **1990**.

[9] J. Brakenhoff, H. T. Van Der Voort, N. Nanninga, *Anal. Chem. Acta* **1984,** *163*, 213.

[10] C. Sheppard, *J. Phys. D* **1986,** *19*, 2077.

[11] DILOR, *Confocal Laser Raman Data Sheet* **1992,** Lille, France.

[12] P. Dhamelincourt in *Microbeam Analysis* (Ed.: D. E. Newbury), San Francisco Press, San Francisco **1979,** p. 155.

[13] P. Dhamelincourt, F. Wallart, M. Leclercq, A. T. Nguyen, D. O. Landon, *Anal. Chem.* **1979,** *51*, 414A.

[14] K. P. J. Williams, G. D. Pitt, B. J. E. Smith, A. Whitley, D. N. Batchelder, I. P. Hayward, *J. Raman Spectrosc.* **1994,** *25*, 131.

[15] K. K. Liu, L. H. Chen, R. S. Cheng, M. D. Morris, *Appl. Spectrosc.* **1991,** *45*, 1717.

[16] A. Govil, D. M. Pallister, L. H. Chen, M. D. Morris, *Appl. Spectrosc.* **1991,** *45*, 1604.

[17] A. Govil, D. M. Pallister, M. D. Morris, *Appl. Spectrosc.* **1993,** *47*, 75.

[18] J. Barbillat, Doctoral Thesis **1983,** Université de Lille I, France.

[19] D. J. Gardiner, C. J. Littleton, M. Bowden, *Appl. Spectrosc.* **1988,** *42*, 15.

[20] M. Bowden, D. J. Gardiner, G. Rice, *J. Raman Spectrosc.* **1990,** *21*, 37.

[21] J. Barbillat, M. Delhaye, P. Dhamelincourt in *Microbeam Analysis* (Ed.: G. W. Bailey, J. Bentley and J. A. Small), San Francisco Press, San Francisco **1992,** p. 1514.

[22] P. Dhamelincourt, J. Barbillat, M. Delhaye, *Spectrosc. Eur.* **1993,** *5*, 16.

[23] J. Barbillat, P. Dhamelincourt, M. Delhaye, E. Da Silva, *J. Raman Spectrosc.* **1994,** *25*, 3.

[24] J. Barbillat, E. Da Silva, J. L. Hallaert, *J. Raman Spectrosc.* **1993,** *24*, 53.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

4 Three-Dimensional Light Microscopy

4.1 Introduction

Real objects are not flat but have a thickness and a surface height variation; that is, a topology. Although conventional light microscopes can be used to study different layers in an object and to measure the differences in height, for some reason these features were not taken account of in a rigorous manner until about 10 years ago. The merging of computers, cameras, sophisticated electronics, and the light microscope has introduced a new attitude and provided us with the tools to exploit fully the capabilities of the nondestructive interaction of light and matter.

The most important developments in the past few years have been the advent of confocal microscopy and special software for deconvolving stacks of images. In particular, the confocal microscope has made three-dimensional microscopy a permanent and lasting tool which is available in many modern laboratories. The large number of users has, in turn, resulted in an improved understanding of three-dimensional image formation in light microscopy.

This Chapter covers some of the issues of importance in the material sciences. It emphasizes the importance of telecentricity to the whole concept of three-dimensional microscopy, introduces the most important aspects of three-dimensional image formation, and outlines problems that provide fundamental limits to the way in which three-dimensional microscopes are used today. The final paragraphs emphasize developments that might result in higher resolution along all three axes.

4.2 Thin and Thick Objects

The theory of light microscopy is usually taught with flat specimens in mind. Especially the well-known limit to the resolution Δx of transmission light microscopes [1]:

$$\Delta x = 1.2\left(\frac{\lambda}{\mathrm{NA}}\right) \wedge \mathrm{NA} = n \sin \alpha \qquad (1)$$

which describes the distance of two point objects in terms of the illumination wavelength λ and the numerical aperture NA (i.e. half the opening angle α of the objective lens). Equation (1) is valid only if the two objects are in focus in the same plane.

A flat object is, by definition, an object in which all features are in focus in a single plane. A thick object has several planes in which its features are in focus. This may seem obvious but, as it will become clearer further below, the features of the

observing system must be taken into account. The depth of focus (i.e. the capability to distinguish objects along the optical axis) depends inversely on the square of the numerical aperture [1].

One must also distinguish between objects of different opacity. Opaque objects have a surface, while translucent objects provide a stack of images in which each layer contributes to the image of the object.

4.3 Telecentricity

Crucial to the understanding of three-dimensional imaging is the principle of telecentricity. All microscopes (in fact all optical systems that measure distances in an object) use telecentric arrangements. The basic idea of telecentricity is that two lenses are used to form an image and these two lenses share a common focal plane (Fig. 1). In a basic Keplerian arrangement this common focal plane is between the lenses. An optical stop (a diaphragm) is centered around the common focus on the optical axis. This stop defines the beam diameter and, together with the focal length, the numerical aperture of the optical system, and therefore also its resolution.

In a telecentric system, the lateral and axial magnifications are independent of the position of the object along the optical axis (Fig. 2):

$$x_{\text{img}} = M \cdot x_{\text{obj}}$$

$$z_{\text{img}} = M^2 \cdot z_{\text{obj}}$$

$$\tan \alpha_{\text{img}} = M^{-1} \cdot \tan \alpha_{\text{obj}}$$

$$M = \frac{f_2}{f_1} \qquad (2)$$

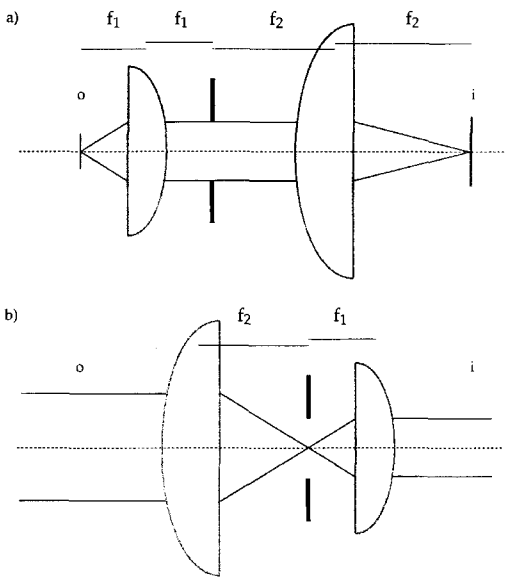


Figure 1. A telecentric system requires two lenses. The distance of the two lenses is the sum of their focal lengths $f_1 + f_2$. o and i indicate the object and its image, respectively. (a) The common focal plane is the position of the diaphragm, the diameter of which, together with the focal length, defines the numerical aperture of a microscope objective lens. The magnification of a telecentric system is the ratio of the focal lengths of the two lenses. (b) A telescope uses an essentially identical arrangement. The object is, however, located at infinity. A telecentric system is, therefore, by definition afocal.

If an object is not in focus and its boundaries are not clearly visible, its size will not change. The light distribution of a single point object will spread as the sample is moved away from the focal plane, but the center–center distance of two point objects will not change. This distance is a function only of the center–center distance in the object and the magnification of the optical system.

A microscope objective lens must therefore be regarded as a compound lens consisting of at least two simple lenses with focal lengths f_1 and f_2. If the magnification of a lens is 100× and the image is 160 mm

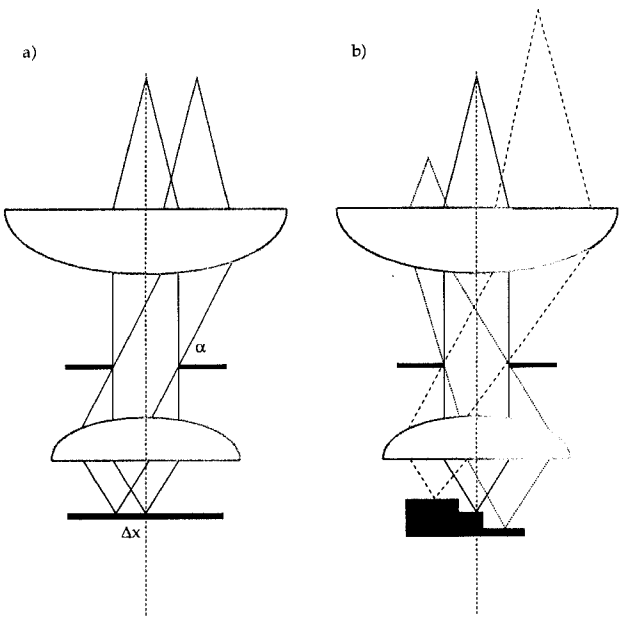


Figure 2. (a) In a telecentric optical system all beams stemming from the focal plane pass the diaphragm as a plane wave. The angle at which the beam passes the diaphragm depends on the distance of the point from the optical axis from which the beam stems. Each position in the focal plane hence has a unique pair of associated angles. (b) In three dimensions, the situation becomes slightly more complicated. The lateral distances are independent of the position of the emitter along the optical axis, but the images of the emitters are spread along the optical axis. The lateral positions are therefore still encoded by a pair of angles, and a divergence or convergence angle defines the position of the emitter along the optical axis.

away from the lens, it is reasonable to assume that the focal length of the smaller lens (which is closer to the object) is about 1.6 mm. When using an immersion system the ratio of the respective refractive indices has to be taken into account. The infinity-correction optics are no exception. The tube lens probably has a focal length of around 160 mm and shares a common focal plane with the microscope objective lens corrected for infinity.

The telecentric plane is also referred to as the 'backfocal plane' of the microscope objective lens. If an object is in the focal plane of a lens, its Fourier-transformed image is found in its conjugate focal plane or backfocal plane. The second lens will Fourier transform the Fourier transform and hence form a real image in its backfocal plane, which in a microscope is a conjugated image plane. The important property of the Fourier transform is that every position in the object has an associated pair of angles in a plane wave passing the backfocal plane.

4.4 Theory of Three-Dimensional Imaging

In light microscopy, resolution is defined by the extent of the point spread function (PSF) which is a mathematical description of the intensity distribution of a point source in the focal region. The smaller the extent of the PSF of the microscope objective lens, the better the distinction between separate points and hence the resolution of the microscope. To summarize the detailed theoretical investigations by many authors, an image can be calculated from the disturbance of the electric field caused by the object [1]. A function that is proportional to this disturbance function is convoluted with the amplitude

PSF $K(x_1 - x_0, y_1 - y_0)$:

$$U_1(x_1, y_1) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} U_0(x_0, y_0) \cdot K(x_1 - x_0, y_1 - y_0)\, dx_0\, dy_0 \quad (3)$$

The theory makes two assumptions: invariance and linearity. Invariance essentially means that a function exists which describes the image-formation process for every point in the object. Linearity means that the image formation is independent of the object. Although this is not obvious at first sight, the principle can be extended to the three-dimensional case. The integral (Eq. (3)) is then performed in three dimensions instead of two, and it is the three-dimensional disturbance of an electric field that has to be calculated.

The integral can be abbreviated using the convolution operator:

$$U_1 = U_0 \otimes K \quad (4)$$

Two cases have to be distinguished [2]. Reflection will, in general, maintain the coherence of the electric field and the intensity I_1 is derived by taking the modulus squared of U_1:

$$I_1^{\text{refl}} = U_1 \cdot U_1^* = |U_1|^2 = |U_0 \otimes K|^2 \quad (5)$$

Fluorescence, on the other hand, does not maintain the coherence. Fluorescence emission is proportional to the intensity and not to the amplitude. The emission is therefore calculated by convoluting the modulus square of the PSF and the modulus square of the electric field disturbance due to the object:

$$I_1^{\text{fl}} = |U_0|^2 \otimes |K|^2 \quad (6)$$

This has a number of interesting consequences which we will only be able to touch on in this context.

4.5 Confocal Reflection Microscopy

Confocal microscopy plays an important role in three-dimensional imaging of thick samples because of its depth-discrimination capability (Fig. 3). This is due to the combination of a point illumination and

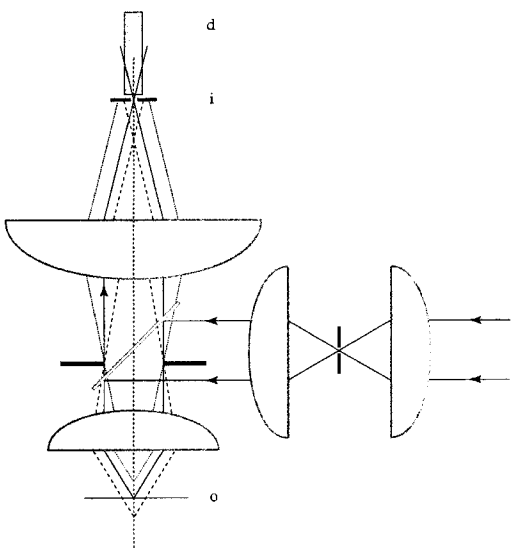


Figure 3. The principal layout of a confocal microscope. A laser beam is focused on an illumination pinhole, collimated, and deflected towards a microscope objective lens which focuses the light on the object. Light stemming from the focal plane is collected by the lens, passes the deflector (e.g. a 50% or dichroic mirror) as a plane wave and is focused on a pinhole located in a conjugate image plane. The photon-sensitive detector d converts the photon flux into a current which can be measured as a function of the beam's position inside the object. Points in front of the focal plane have an image behind the detection pinhole (dotted line), while points that are behind the focal plane have an image in front of the detection pinhole (dashed line). These beams are expanded in the conjugate image plane and, therefore, the detection pinhole will discriminate against the light stemming from planes other than the focal plane. The discriminating effect is lost when the detection pinhole is removed and a detector with a large sensitive area is used.

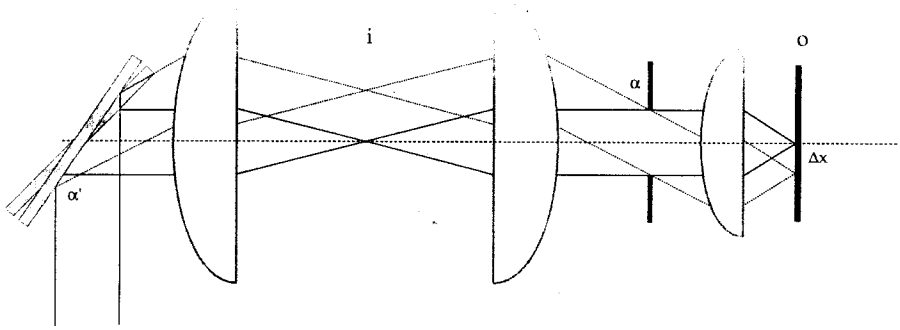


Figure 4. In a confocal beam scanning microscope, one or two mirrors located in planes conjugate to the diaphragm plane of the telecentric system are tilted and thus deviate the beam by an angle α' which is proportional to α. This causes a light spot to move in the image plane i. The mirrors are located in a plane that is conjugate to the backfocal plane of the objective lens. Galvanometer mirrors are the most common devices to tilt the beam, but alternatives such as acousto-optical modulators, resonant galvanometers, and rotating polygons have specific advantages.

the restriction of detection by a pinhole in front of a light-sensitive detector [3]. The depth discrimination is usually interpreted as an improved axial resolution. The improved lateral resolution is of no practical importance.

A system such as the one shown in Fig. 3 does not produce an image. It is a sampling device. Signals must therefore be produced as a function of the beam's or object's position relative to the position of the optical axis and the focal plane. By moving the beam and/or the object three-dimensional confocal intensity distributions are acquired (Figs. 4 and 5).

4.5.1 Position of the Scanning System

The correct position of the scanning system is obviously crucial for the performance of a scanning laser microscope [4]. The scanner performs two functions: (1) The initially stationary illumination beam is tilted and hence the focus is moved laterally in the focal plane (in two dimensions; Fig. 4); and (2) the reflected light is deflected towards the detection pinhole. Since there is practically no time delay (reflection is instantaneous and the fluorescence decay half-life is of the order of nanoseconds), incoming and outcoming light will always follow the same optical path. The two important positions in a microscope are the image plane and the backfocal or telecentric plane. The image of a single point in an object is given by

$$\mathcal{F}: F(\omega) = \int_{x=-\infty}^{x=\infty} f(x) \exp\{i\omega x\}\, \mathrm{d}x \qquad (7)$$

$f(x)$ is the pupil function which is usually equal to 1 in a disk and 0 (i.e. nontransparent) outside the disk; the integral describes the Fourier transform of the pupil function. What the equation also says is that each position in the object plane can be associated with a pair of angles ϕ, θ in the backfocal plane and vice versa.

$$\begin{aligned} x, y &\xrightarrow{\mathcal{F}} \phi, \theta \\ \phi, \theta &\xrightarrow{\mathcal{F}^{-1}} x, y \end{aligned} \qquad (8)$$

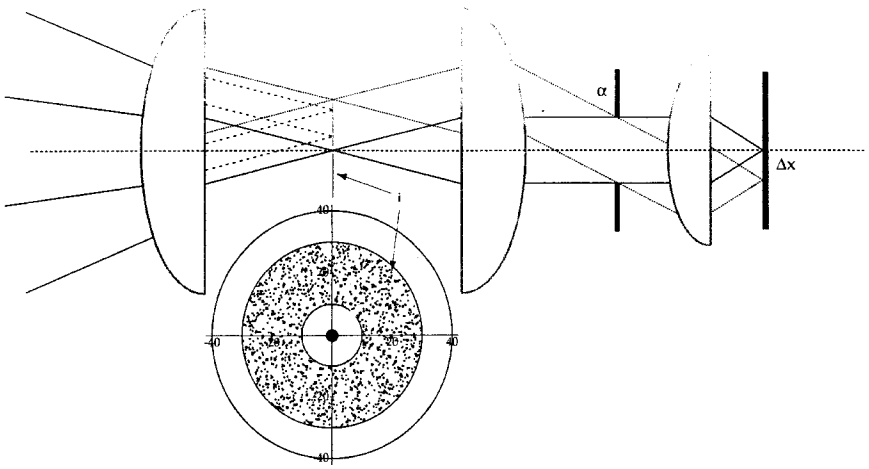


Figure 5. An alternative to beam scanning is the rotating disk. A white light source illuminates a disk which is located in a plane conjugate to an image plane i of a microscope. The disk is scattered with holes with a minimal distance. Most of the light will be reflected by the disk but some of it will pass the holes and hit the specimen. The reflected/emitted light from the specimen will pass the same holes and can be separated from the illumination light by standard optical techniques. By rotating the disk an appropriate pattern eventually covers the whole object and a confocal image is formed in a camera or in the eye of the observer (so-called 'direct view'). Rotating disk, moving slits, and other similar devices are especially useful in the materials sciences when the low emission/reflected intensity can be compensated by a brighter light source.

By placing the scan mirrors in planes that are conjugate to the backfocal plane, and by tilting them by ϕ, θ in orthogonal axes, a light spot can be moved to any position in the object plane. The light beam that is deflected by the two mirrors must be collimated.

In an ideal system both mirrors will be placed in planes that are conjugate to the telecentric plane. It is usually sufficient to have both mirrors very close to the telecentric plane and it is then only required to have both scan mirrors very close to each other with the telecentric plane between them. An alternative is to tilt a single mirror in two axes [4].

4.5.2 Resolution

Ultimately the resolution of any optical instrument is determined through its contrast transfer function. Also if resolution is specified as the full-width half-maximum (FWHM) of the PSF, it is the contrast that determines the performance of an instrument. The *lateral resolution* can be determined by measuring the size of small particles or by measuring distances between two particles or two lines. An estimate for the point resolution Δx in terms of the FWHM is [5]:

$$\Delta x = \frac{1.2}{\sqrt{2}} \cdot \frac{\lambda}{\mathrm{NA}} \tag{9}$$

The point resolution can thus be expected to be about 40% better than in a conventional microscope (Eq. (1)) if two flat samples are compared.

The *axial resolution* is much more complicated to measure (Figs. 6, 7 and 8). In principle, a single point could be observed in a plane parallel to the optical axis and

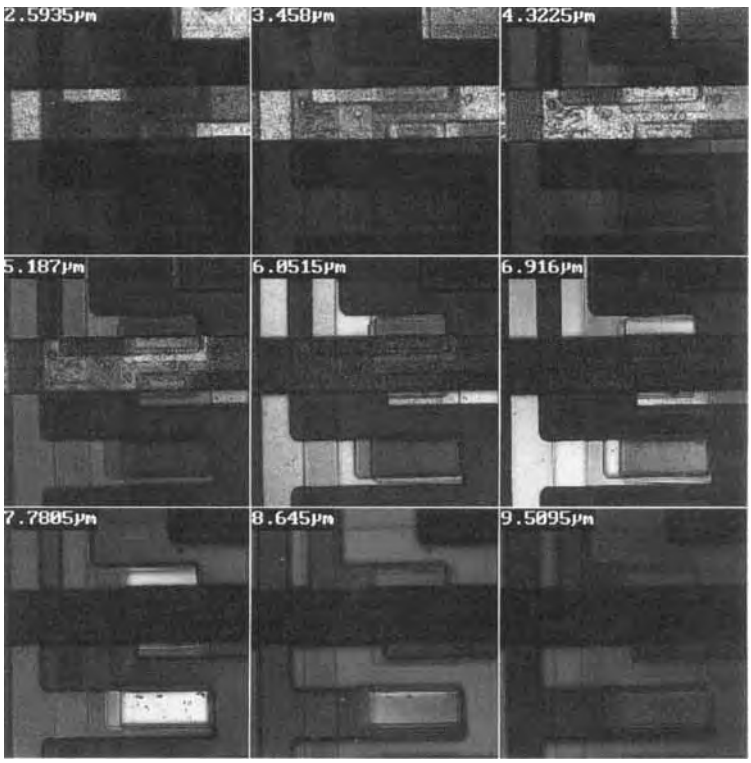


Figure 6. A confocal reflection through-focus series of images of an integrated circuit recorded using a Carl Zeiss LSM 410 (illumination wavelength 543 nm, oil-immersion lens 63×/1.25, field size 88 μm). The series covers a range of approximately 7 μm. The axial distance between two images is approximately 0.9 μm. Clearly visible is how various parts of the integrated circuit come in and go out of focus and merely change their intensity.

both resolutions could be observed in a single x/z image. An estimate for the axial resolution Δz for point objects in terms of the FWHM is [6]:

$$\Delta z_{\mathrm{fl}} = 1.5 \cdot n \cdot \frac{\lambda}{\mathrm{NA}^2}$$
$$\Delta z_{\mathrm{ref}} = 1.0 \cdot n \cdot \frac{\lambda}{\mathrm{NA}^2} \qquad (10)$$

The axial resolution for a reflecting object is thus about 30% better than for a fluorescing object. Note that the resolution decreases with an increasing refractive index n of the immersion system. This means that an air lens with a numerical aperture of 0.95 has a higher axial resolution than an oil immersion lens with a numerical aperture of 1.4. The factor of 1.5 is true only for the observation of FITC at a wavelength of 488 nm and a detection around 590 nm. It depends on the ratio of the excitation and emission wavelengths and will increase when fluorophores with a large Stokes shift are observed, but the estimate is probably sufficient for practical purposes.

Figure 9 illustrates that resolution should not be confused with measuring the height of a plane over the height of another plane. Since now two planes have to be distinguished from one another, in principle it is possible to measure a relative height with an essentially unlimited resolution.

4.5.3 Interference

A serious problem is the interference from layers above and below the plane of focus.

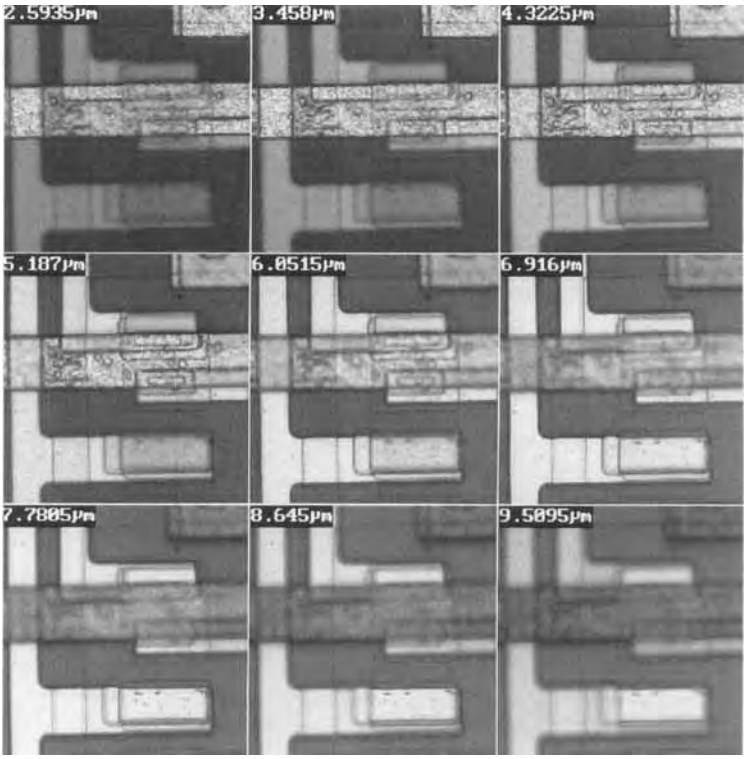


Figure 7. The same area as in Fig. 6 observed in nonconfocal mode. The whole object is always visible, indicating that light discrimination is much more significant in confocal than nonconfocal contrasts. Structures that are out of focus are slightly dimmer and the features appear to be fuzzy. To generate this image the pinhole was removed.

This is the case, for example, in the slightly transparent silicon oxide layers which cause multiple images and axial interference fringes. A solution, which has been explored by Hell et al. [7], is incoherent illumination using white light. There are also several claims that interference can be used to improve the lateral and axial resolution [8, 9], but this has not resulted in an impact on commercial instruments.

4.6 Confocal Fluorescence Microscopy

As pointed out above, imaging theories essentially assume that the process of image formation is linear. This means that all the energy is conserved or that all the light which enters an object will be collected. In three dimensions this means that all the light which is present in one layer is also present in another layer. This is obviously in conflict with many useful contrasts. Fluorescence can only be induced when the fluorophore absorbs light. Thick objects are therefore unlikely to receive the same amount of light in every plane. Worse still, the amount of light received in one plane depends on the amount absorbed during the penetration process. Therefore, the image recorded inside a thick specimen is, in principle, nonlinearly dependent on the object. The direct consequence is that the theory as it has been presented above does

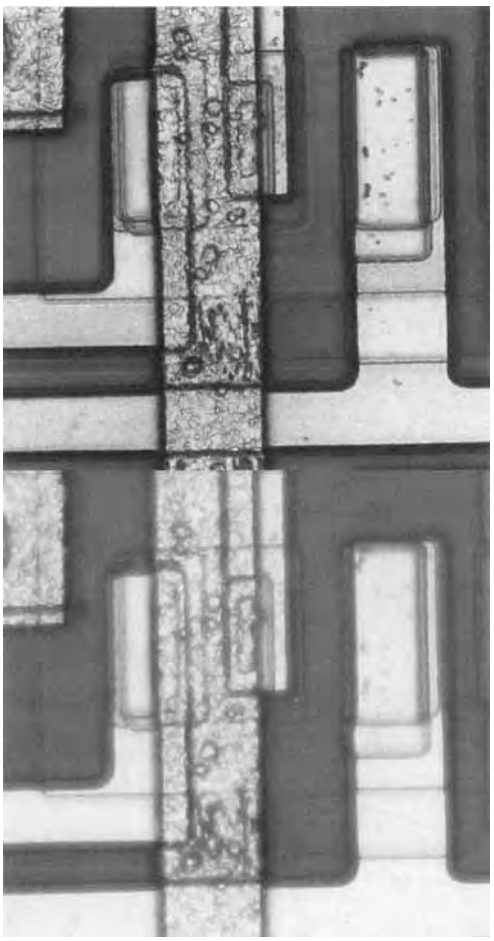


Figure 8. The two images were obtained by selecting the maximum intensity along the vertical axis for each picture element, thus discarding all low intensities (so-called 'extended view'). The confocal image on the right-hand side is obviously much sharper, which in this case really means that the contrast is better and features are more easily distinguishable.

not hold, and a transfer function does not exist.

Contrasts such as fluorescence can therefore only be applied to thin objects and objects which have a low absorbance. As a general rule, specimens that have an absorbance which is clearly visible to the naked eye are unsuitable for a quantitative evaluation. Many biological objects that are treated with absorbing dyes, polymers, etc., fall into this category. Most fluorescent-antibody-labeled specimens are transparent in a conventional transmission contrast and more likely to provide reasonable images.

4.7 New Developments in Fluorescence Microscopy

4.7.1 Multiphoton Illumination

The axial and lateral resolutions depend only on the numerical aperture of the optical system and the illumination and emission wavelengths. To improve further an instrument's resolution axially or laterally or along all three axes, there are thus two obvious possibilities: the numerical aperture of the system can be increased; and the wavelengths can be decreased. Decreasing the wavelengths is, however, limited by the optical properties of glass. The use of ultraviolet (UV) illumination can be avoided by applying two-photon absorption [10], but the detection of the fluorescence emission is still affected. Besides other advantages over single-photon absorption, a confocal two-photon fluorescence microscope can exhibit the resolution of a confocal fluorescence microscope operating in the UV far below 360 nm [11].

4.7.2 Multiple Lenses

Recent developments of new instruments use two or more objective lenses to observe

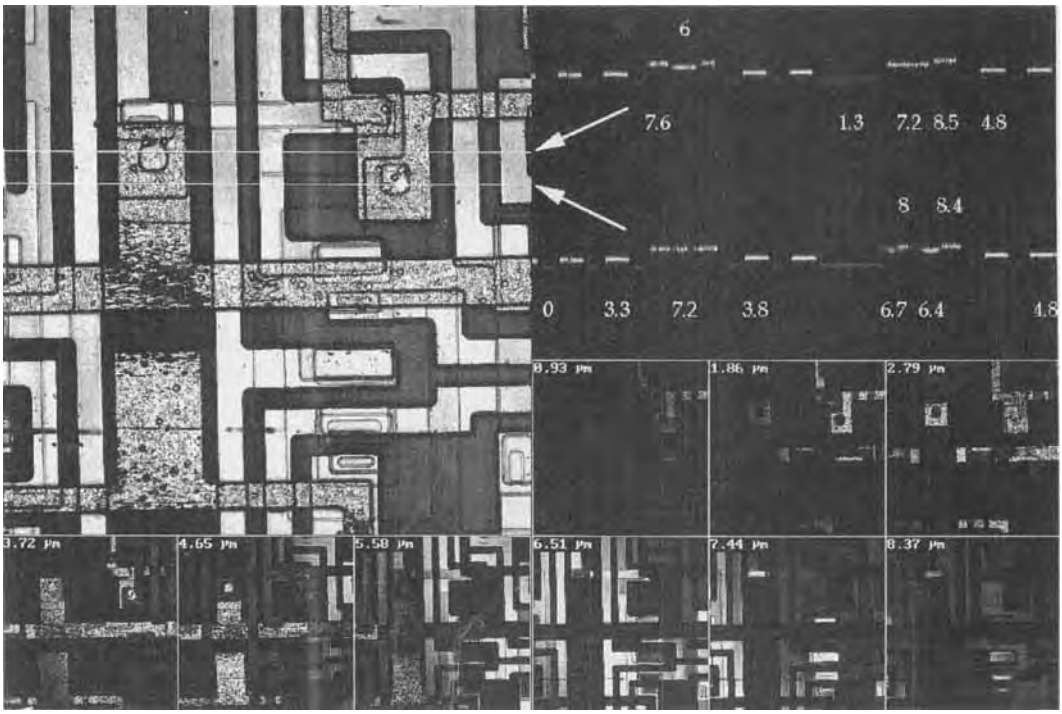


Figure 9. Images of an integrated circuit. (a) An extended view in which two lines indicate where x/z images were recorded (field size 210 μm). (b) The two x/z images pass through different lines in the integrated circuit. The numbers indicate the height above normal in which certain features can be found. The precision is in this case limited by the number of lines that were used to acquire the image. (c) The series of images along the optical axis used to generate the extended view.

the sample. Thereby the light distribution in the focal plane can be changed (e.g. by interference [12, 13]), or the spatial arrangement of the illumination and detection PSFs in a confocal microscope can be changed; that is, the illumination PSF can be shifted or rotated relative to the detection PSF [14].

In 4Pi-confocal fluorescence microscopy [12], the sample is coherently illuminated and/or observed through two opposing objective lenses. The light interferes in the sample and/or in the detector and leads to a substructured PSF with several maxima along the optical axis. The axial FWHM of the central peak in the 4Pi-confocal PSF is much smaller than the FWHM of the confocal PSF, and the observation volume is effectively reduced by a factor of 2. This improvement has been shown experimentally for fluorescence and for scattered light microscopes [15]. A similar, nonconfocal, arrangement that makes use of interference to increase the axial resolution is standing-wave fluorescence microscopy [13].

Even in a high numerical-aperture objective lens, less than 35% of the full solid angle of 4π are covered in the illumination and detection processes. The lateral resolution of a confocal microscope is,

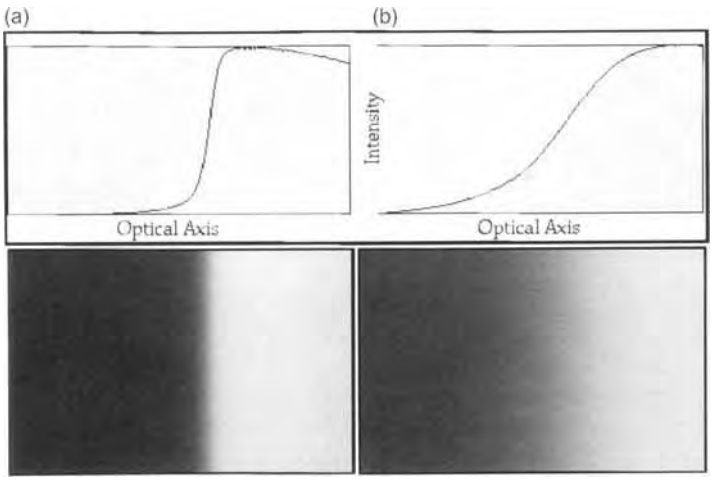


Figure 10. Demonstration of axial discrimination in confocal fluorescence microscopy (illumination wavelength 543 nm; detection, above 570 nm; axial resolution, approximately 700 nm (20%–80%)). Rhodamine 6G was dissolved in immersion oil and mounted between a cover slip and a microscope slide. Images were then recorded along the optical axis, moving from the cover slip on the left-hand side to the fluorophore. The depth discrimination of the confocal arrangement (a) results in a steep slope while the removal of the pinhole in front of the detector (b) results in a lower discrimination along the optical axis. The slope can be analyzed and provides a convenient way to determine the resolution of a confocal fluorescence image. About 6 times more light was collected in the nonconfocal image.

therefore, at least a factor of 2 better than the axial resolution. In consequence, the observation volume is an ellipsoid elongated along the optical axis. Recently, a new microscopic set-up that exploits this effect was proposed [14]. Two objective lenses are used to illuminate the sample and to collect its fluorescence emission at an angle to the illumination axis. The resolution enhancement stems from the alignment of the lenses: the detection axis is approximately orthogonal to the illumination axis. Therefore, the overlap of the illumination and detection PSFs is minimized and the lateral resolution of the objective lens becomes dominant. Thus, the microscope has high axial and lateral resolutions in all directions.

4.8 References

[1] M. Born, E. Wolf, *Principles of Optics*, 6th edn, Pergamon Press, Oxford **1980**.
[2] T. Wilson, C. J. R. Sheppard, *Theory and Practice of Scanning Optical Microscopy*, Academic Press, London **1984**.
[3] C. J. R. Sheppard, A. Choudhury, *Opt. Acta* **1977,** *24*, 1051.
[4] E. H. K. Stelzer in *Handbook of Biological Confocal Microscopy* (Ed.: J. B. Pawley), Plenum Press, New York **1995,** 139–154.
[5] G. J. Brakenhoff, P. Blom, P. Barends, *J. Microsc.* **1979,** *117*, 219.
[6] R. Müller, Diplomarbeit, Ruprecht-Karls-Universität, Heidelberg **1990**.
[7] S. Hell, S. Witting. M. v. Schickfus, R. W. Wijnaendts van Resandt, S. Hunklinger, E. Smolka, M. Neiger, *J. Microsc.* **1991,** *163*, 179.
[8] T. Wilson, R. Juškaitis, *Bioimaging* **1994,** *2*, 36.
[9] R. Juškaitis, T. Wilson, *J. Microsc.* **1995,** *176*, 188.

[10] W. Denk, D. W. Piston, W. W. Webb in *Handbook of Biological Confocal Microscopy* (Ed.: J. B. Pawley), Plenum Press, New York **1995,** 445–458.

[11] E. H. K. Stelzer, S. Hell, S. Lindek, R. Stricker, R. Pick, C. Storz, G. Ritter, N. Salmon, *Opt. Commun.* **1994,** *104*, 223.

[12] S. Hell, E. H. K. Stelzer, *J. Opt. Soc. Am. A* **1992,** *9*, 2159.

[13] B. Bailey, D. L. Farkas, D. L. Taylor, F. Lanni, *Nature* **1993,** *366*, 44.

[14] E. H. K. Stelzer, S. Lindek, *Opt. Commun.* **1994,** *111*, 536.

[15] S. Lindek, E. H. K. Stelzer, S. W. Hell in *Handbook of Biological Confocal Microscopy* (Ed.: J. B. Pawley), Plenum Press, New York **1995,** 417–430.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

5 Near Field Optical Microscopy

5.1 Introduction

Near field microscopy was born in the early 1980s in the IBM–Zürich Research Laboratory [1]. For a long time other laboratories had worked on the old problem of exceeding the diffraction limit in classical imaging systems [2–4]. It is known that diffraction causes a disastrous effect on the spatial resolution, which cannot be better than $2NA/\lambda$, where NA is the numerical aperture of the microscope and λ the wavelength of the illuminating radiation [5]. Assuming that the angular aperture is 0.95 (this corresponds to an excellent microscope objective) the smallest periodic structure which could be imaged would be about 350 nm (for a mean wavelength of 550 nm). In the 1960s and 1970s a few techniques for increasing the resolution of classical microscopes were developed [6, 7]. These ranged from synthetic apertures to numerical superresolution. Unfortunately, neither optical synthesis nor analytical techniques led to a substantial improvement in resolution. Finally, in the 1960s, the invention of the scanning optical microscope allowed an increase in the resolution by a factor of $\sqrt{2}$ [8].

5.2 Principle of Near Field Microscopy

Basically, a near field microscope is a scanning microscope whose objective has been replaced by a small needle, the role of which is to emit, collect or perturb the light over the object surface. This peculiar emitter/collector can be either in a dielectric, a semiconducting material, a metal or even a dielectric metal-coated material. A simple near field microscope is described in Fig. 1a. Among the main elements is a scanning system, electronically driven and computer controlled. This system, which enables the sample to be scanned, is generally based on the particular properties of a piezo tube in a PZT ($PbZrO_3$–$PbTiO_3$) ceramic. This tube is metallized on its outer and inner sides. The outer side is then divided into four strictly identical parts playing the role of electrodes (Fig. 1b) [9]. Note that such a device is used in many other local probe microscopes such as STM or AFM microscopes. Its advantage is simplicity, although it undoubtedly has a small scanning area. As an example, a piezo tube of 6 mm (outer diameter), 0.8 mm (thickness) and 6 cm (length) provides a scanning area of about 30 μm when voltages of ±80 V are applied on two opposed electrodes. A second main

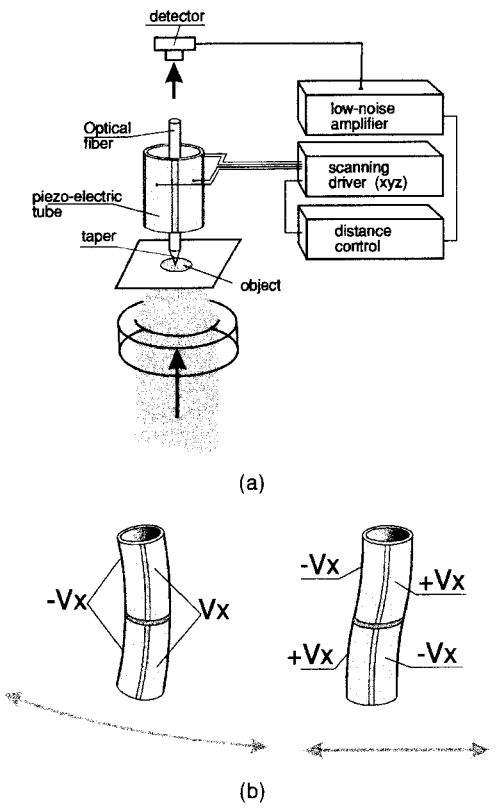


Figure 1. (a) Basic scheme of a near field microscope (working here in the collection mode). The small needle collects the light on the object surface. The image is generated by scanning the surface with an $x-y$ piezo translator. (b) By a suitable electrode segmenting it is possible to compensate the circular deflection of the piezo tube.

element is the tip itself, which is described in the following.

5.3 Basic Families of Near Field Microscopes

5.3.1 Background

To analyze the properties of the different configurations of near field microscopes it seems useful to classify them into families characterized by common features. Chronologically we distinguish three kinds of nanocollectors or nanoemitters: the first is a small hole pierced in an opaque screen (proposed by Synge in 1928 [2]), the second is based on the use of a taper in a dielectric material (proposed by Pohl et al. in the 1980s [1]). The latest comprises nano-scattering components, which can be either transparent or opaque. One of the first was proposed by Wickramasinghe et al. [10]. O'Keefe [3], Ash and Nicholls [4], Massey [11], and Fischer and Zingsheim [12] proposed nanocollectors and -emitters belonging to the first class. O'Keefe's configuration was never realized [3]. That of Ash and Nicholls can be employed with microwave radiation [4], while the configuration proposed by Massey can be used for centimeter and infrared radiation [11], and that of Fischer and Zingsheim can be used in the visible range. Near Field Microscopes based on the use of a taper in a dielectric material were proposed by a group at Cornell [13, 14], Courjon et al. [15, 16], Reddick et al. [17], de Fornel et al. [18], Van Hulst et al. [19], Bozhevolnyi et al. [20] and Bielefeldt [21].

5.3.2 Detection Techniques in Near Field Microscopy

The main problem of detection in every near field microscope is either the low signal or the low signal-to-noise ratio. Consequently, the detector is generally a photomultiplier or a very low noise photocell or an avalanche photocell. To reduce the noise this element can be cooled by means of a Peltier element, a cooling liquid such as

nitrogen, or simply by putting the detector in a freezer. −20 °C is easily reached at a reasonable cost. The signal-to-noise ratio can be improved by a factor of 50 (in intensity). As explained previously, three possibilities are offered to interact with the field: collection of the near field on the object (collection mode); local illumination of the object (emission mode); and, finally, perturbation of the near field (perturbative mode). The detection procedure will thus depend on the type of interaction.

5.3.3 Collection Mechanism in Near Field Microscopy

Although the collection mechanism in confocal microscopy is well understood, this is not the case in near field microscopy. At the outset we must separate the role and the action of the various parts of the nanosensor. First consider the case of collection. Figure 2a shows the scheme of a nanocollector. The light collector is essentially a taper, the shape of which can vary from the simple conical shape to a more sophisticated one called a compound concentrator, the role of which will be described more precisely. Analysis of the tip reveals two distinct parts: the upper suprawavelength and the lower subwavelength. The light must first be collected by the subwavelength extreme tip and then transmitted inside the tip as far as the detector.

5.3.3.1 Role of the Subwavelength Extremity of the Collector

The subject of the collection–emission mechanism has fired much discussion and controversy in recent times. The general feeling is that the subwavelength part of the tip is excited by the light energy lying on the surface as any scatterer. The apex of the tip will simply reemit light which will be partly transmitted inside the upper part. This approach is in agreement with the experimental techniques developed by Courjon et al. [15], Bozhevolnyi et al. [20], Bielefeldt et al. [21], where the nanocollector is a simple bare taper-shaped tip

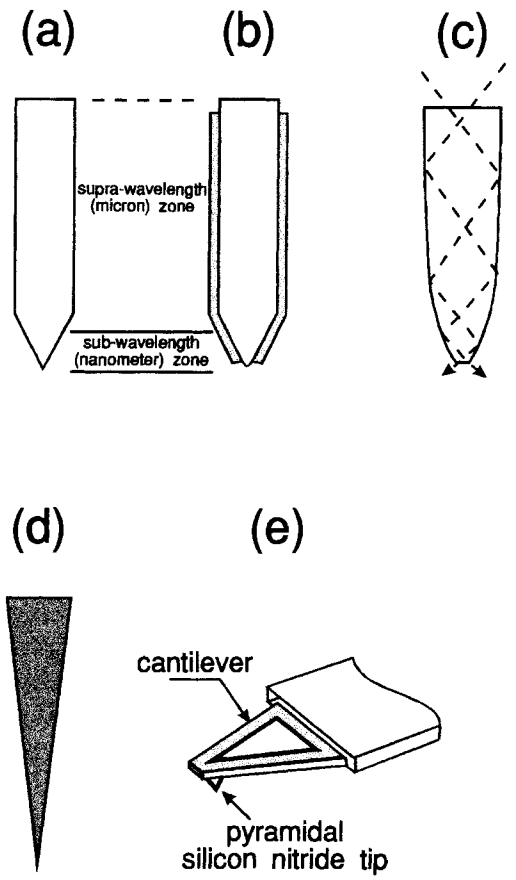


Figure 2. Scheme of the tip. (a) Tip is a simple bare cone-shaped fiber tip, (b) a metallized tip whose extremity is free from any metal, (c) a paraboloidal shaped tip, (d) a tip used in the perturbation mode, and (e) a silicon nitride tip associated to a cantilever.

[22]. However, this less obviously explains the working of nano-aperture based techniques with scattering/reemission notions [1]. In this case the nano-collector is a metallized cone-shaped fiber whose very tip is kept free of metal. The obvious aim of this approach is to limit the detecting zone to that of the transparent part of the tip (Fig. 2b) [1, 14]. In fact, this approach has a few advantages, the metallized walls limit the stray light and the metal can enhance the detection capability of the tip [23].

5.3.3.2 Role and Efficiency of the Propagating Part of the Collector

Let us consider Fig. 2a and 2b once more. The macroscopic part is merely a kind of micrometer funnel, the role of which is to guide the light beam to the detector. To obtain the maximum of intensity at the very tip, some investigations have been carried out to best funnel the light beam in this macroscopic part of the tip. In the specific case of the collection mode, the aim is to obtain the maximum amount of light density at the apex level, regardless of the astigmatism of the collected light. The problem of best collecting a light beam, whatever the astigmatism, was studied extensively some years ago by Winston in the context of a study on solar energy concentration [24]. He demonstrated that a suitable collecting surface derived from conicals is able to focus a light beam with excellent efficiency. Such systems, called 'compound collectors', have a typical pseudo-paraboloidal shape (Fig. 2c). This peculiar system [25] is able to preserve the intensity and to ensure a maximum of intensity at the very tip. It is clear that such a method cannot be applied easily in the microscopic range. However, the most efficient tips used by some workers seem to exhibit such pseudo-paraboloidal profiles.

For the limits of applicability of this focusing system we have born in mind that compound concentrators work according to the principles of geometrical optics. Consequently, the limits of validity should be around a few micrometers. However, experiments have shown that this geometrical approach can often be pushed as far as 1 μm. Thus we may reasonably assume that the compound concentrator will be efficient until its diameter decreases to about 1 μm.

5.3.4 Emission Mechanism in Near Field Microscopy

According to Babinet's reciprocity theorem, we expect the nanosensor to function whatever the direction of light. In this approach, the light is launched through a taper-shaped piece of transparent material (generally an optical fiber). The light beam, after more or less efficient funneling, is transmitted to the very tip, which will reemit the light playing the role of nanosource (or nanoantenna). Two explanations for the emission mechanism are in competition. On the one hand, the last few nanometers at the very tip always play the role of a simple scattering center. On the other hand, and especially in the case of a metallized tip, the existence of a physical aperture in the metal will create a nanosource. Tip modeling [26, 27] shows that the problem is very complex because the notion of a nanometer hole is meaningless in optics.

After emission, the far field transmitted through the object (or reflected on the object) is collected by means of a large aperture objective, for example a Cassegrain mirror combination or a paraboloidal concentrator. Note that the resolution of this far field collector does not influence the resolution of the image. The only role of the collector is to pick up the maximum number of photons.

5.4 Perturbation Mode

The perturbation mode, which was simply derived from the collection and emission modes was introduced in 1989 [10]. It is based on the fact that the near field lying on the object surface can be locally perturbed by a small scattering center. The latter will be excited and the resulting coupling will modify slightly the field in the vicinity of the scatterer. If the scatterer oscillates the resulting perturbation can be detected by extracting the oscillating component of the scattered field. The prime interest in such a method is that it allows exceedingly sharp and even opaque tips to be used, because they do not transmit the light; only their extreme tip plays the role of a scattering center (see Fig. 2d). Tungsten tips, like those used in STM, may be employed. Moreover, it is possible to benefit from the high polarizability of metals [23].

This principle of detection can be used in conjunction with the STM or AFM mode. A main drawback comes from the small signal-to-noise ratio due to the detection in the free space rather than through a guiding fiber. Some setups are derived from the STOM/PSTM configuration [28, 29]. The novelty which deals with this basic configuration lies in detecting the reflected beam rather than the light scattered by the tip. To increase the signal-to-noise ratio, a spatial filter is added. When the tip (in metal) is brought very close to the sample, the field perturbation will affect both the evanescent field and the reflected field. By measuring the intensity variations of the latter, it is possible to detect the topography of the sample. If the object itself is metallic, the plasmon resonance will dramatically enhance the perturbation. Finally, it is possible, by measuring the electron tunneling effect between tip and sample, to monitor the tip position with high precision. Some valuable, even unsurpassed, resolutions have been obtained [28].

Another device is derived from conventional microscopy. Assume that a classical microscope objective is used to illuminate a sample. For a very thin oscillating metallic tip placed between the objective and the sample, the magnified image of the sample will be composed of a low resolution image locally perturbed by the oscillating tip (necessarily brought close enough to the sample to perturb the near field). The signal produced by the oscillating component of the detected field will be related to the high resolution details in the image [30]. In 1995, the group of Wickramasinghe improved greatly the device by using an interferometric technique for increasing the signal to noise ratio. A resolution of about one nanometer has been demonstrated [31].

5.5 Main Configurations

Following the previous descriptions dealing with the collection/emission techniques,

near field microscopes may be divided into four principal classes:

- transmission microscopes
- reflection microscopes
- internal reflection microscopes
- hybrid microscopes.

A few collection/emission techniques were proposed in the early 1980s [11, 12] but the first significant results date back to 1983 [1]. This microscope working in the collection mode used a cone-shaped quartz rod and metallized in such a way that the very tip was free of metal. This technique is considered to be the ancestor of the modern near field optical microscopes.

5.5.1 Transmission Near Field Microscope

Transmission mode microscopy was undoubtedly the first technique leading to significant results [1, 14]. It can be subdivided into the emission and collection mode techniques.

In the emission mode [1, 14], Fig. 3a, the tip is used as a nanosource as explained previously. The resulting transmitted field is then detected in far field. An interest in this technique is due to the apparently limited area of the illuminated field, thus preventing a parasitic effect. This area is given by the overall field distribution around the tip. The field distribution itself depends on the tip geometry and the tip–sample interaction. This technique has been developed by the group of Cornell University [13, 14] and some variants have been developed and applied to spectroscopy and to luminescent material analysis, etc. [32, 33].

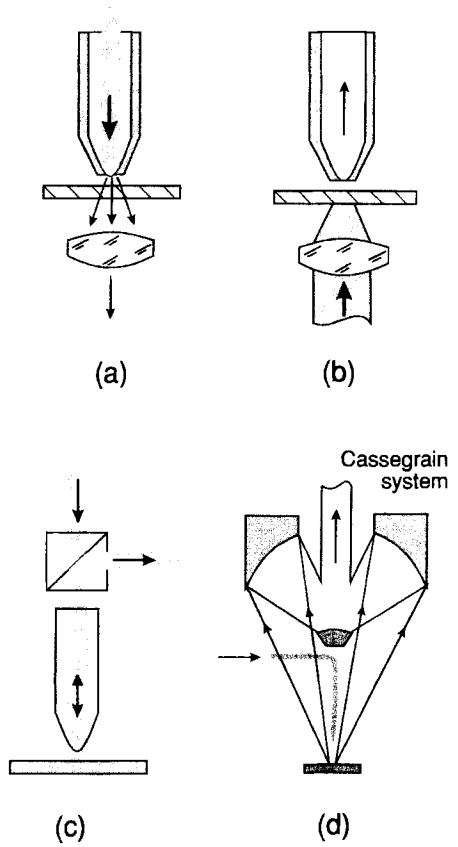


Figure 3. Fundamental microscope configurations: (a) transmission emission mode, (b) transmission collection mode, (c) external reflection mode using the same fiber for emission and collection, (d) external reflection mode using an external collection of the reflected light.

In the collection mode [34], Fig. 3b, the sample is illuminated strongly by means of a well-focused light beam. The illuminated area is thus described by a gaussian-like field distribution, the extension of which is a few tens of micrometers. Since the observed object area is a few micrometers, we can assume that the region to be analyzed is illuminated uniformly. The main consequence of a nonlocal illumination is the risk of nonlocal effects such as long range interference phenomena. For similar

reasons, the emission and collection modes are not equivalent with respect of the polarization effects. In the collection mode it is very easy to polarize the incident light beam, and the collection can be carried out with a classical monomode fiber. The use of a polarization-maintaining fiber as collector will allow analysis of the transmitted light. However, because of nonlocal illumination some spurious interference effects may appear. By contrast, in the emission mode the polarization at the apex is not obvious, although some results exhibit good polarization-maintenance [33]. The transmitted light beam may be analyzed classically by means of an analyzer.

From the literature it seems that the emission mode is preferred to the collection mode. The latter seems to be replaced by the perturbation technique, the principle of which has been described previously.

5.5.2 Reflection Microscopy

The reflection microscope was born almost at the same time as the transmission microscope. The first attempt dates back to Fischer et al. [35], who used a rather sophisticated device. This system used the frustration of an evanescent wave by the protrusions of the object to be analyzed. Some years later the same authors developed a new version of the same setup using a small metallized sphere as an antenna [36]. In this case the object protrusions perturb the plasmon generated in the sphere by the incident beam. Another method, proposed in 1989 [16, 20, 21, 37, 38], consisted of using the same fiber both to illuminate the sample and collect the light reemitted by the object (Fig. 3c). This technique is already used in an operating system although the resolution has not fulfilled yet the hopes of the inventors. Figure 4 shows a test specimen of SiO_2

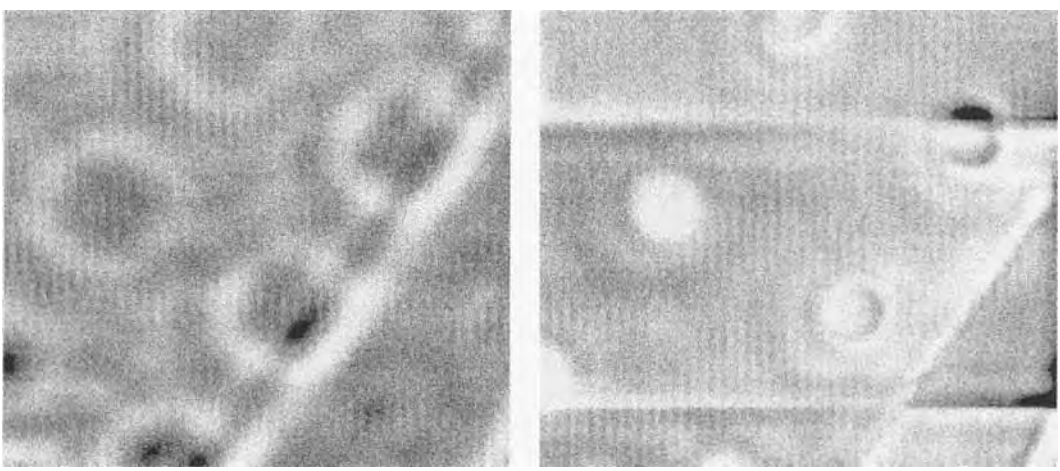


Figure 4. Images in reflection microscopy of SiO_2/Si dots. The thickness is about 100 nm. The two images (the scanned area is 6 μm × 6 μm) are obtained by scanning the object in constant level mode (the tip does not follow the topography). On the left, the tip is about 1 μm distant from the surface. On the right, the tip–sample distance is a few nanometers. The resolution increase in the dot images is dramatic. (The horizontal lines are due to contact between tip and sample.)

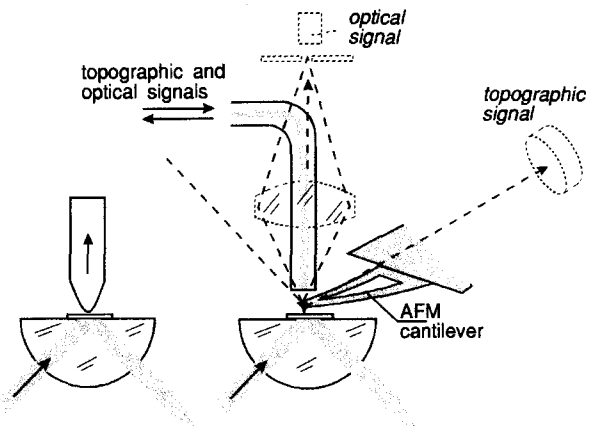


Figure 5. Total internal reflection configuration. (a) Basic setup, (b) combination with atomic force detection. Dashed line: initial setup of this family. The force is detected by a measured deflection of a light beam on the cantilever. The field reemitted by the tip is detected in the far field by means of a conventional objective. Full line: a more compact system where a fiber is used for detecting both the optical field and the topography (AFM interferometric technique).

dots (100 nm thick and about 1 μm in diameter) on a silicon substrate. The tip moves in a given plane (without any distance control system). In Fig. 4a the tip moves at about 1 μm across from the surface, whereas in Fig. 4b the tip is in the near field (a few nanometers distant from the top of the dot). We note the dramatic enhancement of resolution (visible on the dot edges). This result (and also the images in Fig. 8) was obtained with a vertical oscillation associated with synchronous detection to enhance the contrast of the near field contribution.

The third way, proposed by Betzig et al., is a variant of the transmission microscope. Looking at Fig. 3a again we see that it suffices to collect the reflected light by means of a paraboloidal mirror (or one of another suitable shape) to transform a near field transmission microscope into a reflection microscope (Fig. 3d) [39].

5.5.3 Total Internal Reflection Microscopy

The last broad family comprises scanning tunneling optical microscopes (STOM or PSTM). This technique exploits the fact that a light beam impinging onto a prism can be totally internally reflected, thus generating an evanescent plane wave. This internal reflection is used as a peculiar illuminating system, as shown in Fig. 5a. Interest in this approach is due to the fact that the evanescent field is not radiative and does not contaminate the useful signal during the recording. Also, due to the monotonic decrease of the intensity in the evanescent wave, it gives a simple way to control the tip position relative to the object surface. This technique has been used in several ways, such as the direct technique (STOM/PSTM) [15, 17–19], the inverted STOM [40], and the plasmon configuration [28, 41, 42]. Figure 6 shows an image of an almost flat dielectric grating. The resolution (in terms of the smallest visible periodic structure) is about 30 nm.

5.5.4 Hybrid Techniques

One of the main problems related to these basic configurations is the difficulty in clearly defining an analysis plane, as in

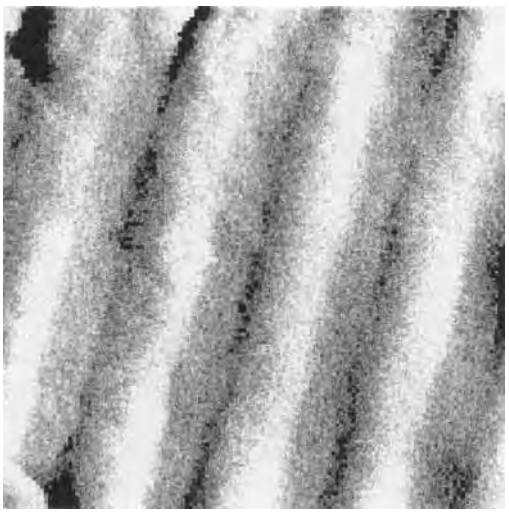


Figure 6. STOM image of a grating with a groove depth of about 5 nm and its period about 400 nm. It is observed in polarized light (TE mode), scanned area 1.4 μm × 1.4 μm.

STM, since the near field generally does not follow the object topography. Trying to use the signal itself to control the tip function is somewhat hazardous. A few attempts have been made to work in the constant intensity mode (similar to the STM constant current mode). This method gives good results for small topography objects [43]. For topography variation larger than a few nanometers, the nonbijective relationship between topography and field prevents the generalized use of the method.

5.5.4.1 Near Field Microscopy by Shear Force Control

To avoid the problem of optical distance control some groups have developed a new external control based on the detection of the force exerted by the sample on the tip. To understand the principle of such control systems, let us assume that the tip vibrates laterally, i.e., parallel to the object surface. When the distance between tip and sample is sufficiently small, the tip–sample interaction damps the oscillation, the amplitude of which can be used as a distance control. Following this principle several configurations have been proposed [44–48] (Fig. 7a, b, c and d). The technique works either in air or in vacuum, even in liquid, whatever the optical configuration. The working distance lies between a few nanometers and a few tens of nanometers. It depends mainly on the interacting surfaces, that is, on the tip surface facing the object. Despite the fact that the interaction has not yet been understood, we must recognize that such a control system is one of the most efficient and the images obtained in this way are excellent. Figure 8a illustrates the capacity of these microscopes to visualize submicron refractive index variations on very flat dielectric objects. This possibility specific to optics is offered neither by STM (working on conducting material) nor by AFM (unable to give information on the physico–chemical properties of the materials).

5.5.4.2 Contact Total Internal Reflection Microscopy

A completely different way of controlling the tip–sample distance consists of putting the tip onto the surface of the object to be analyzed. To avoid or at least to limit the risk of scratches, the tip is mounted on a microcantilever whose stiffness is small enough to ensure a very soft contact. The simplest method then consists of using microcantilevers commercialized for AFM use. With such a technique, the tip follows

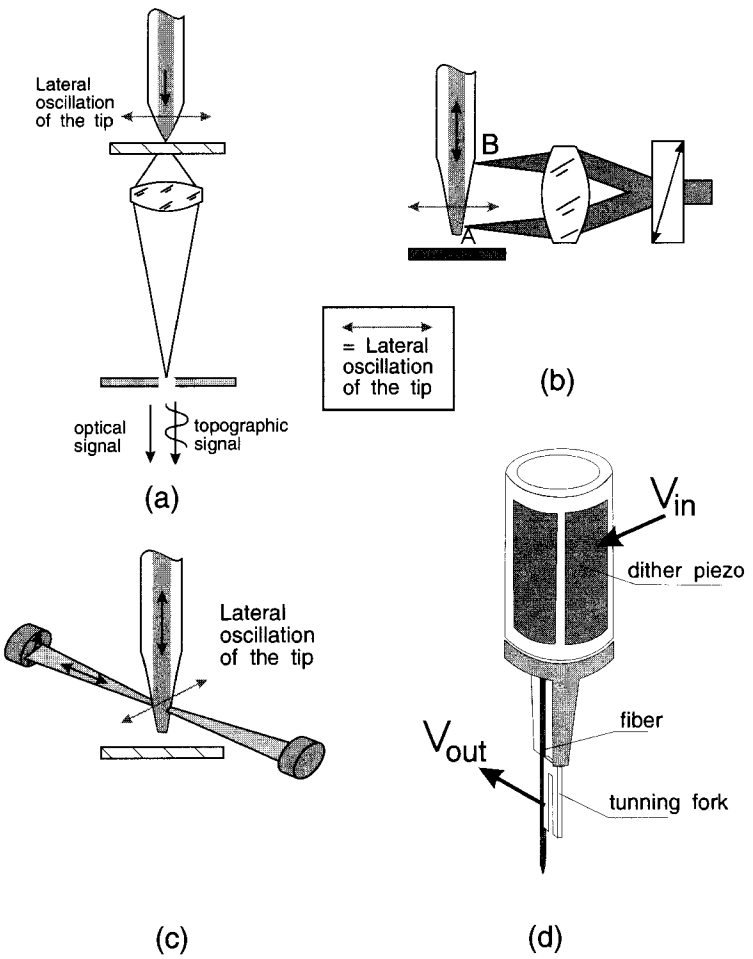


Figure 7. Configurations using a distance control by shear force measurement. (a) Transmission technique, the top oscillates parallel to the object plane. The field transmitted through the object is spatially filtered. The resulting oscillating field component is proportional to the shear force (and thus to the topography) whereas the direct constant component is proportional to the near field. (b) Configuration using interferometry to determine the oscillation amplitude of the fiber. (c) A widely used and very simple way to control the shear force. (d) A non-optical detection using the properties of piezo-electric materials. V_{in} is the periodic oscillation voltage applied to the dither piezo. V_{out} is the signal detected by the electrodes of the tuning fork. Its amplitude and phase depend on the tip-sample coupling.

the true topography and gives information on the field over the surface. At the same time, it is possible to obtain the topography of the sample [49, 50]. The contact technique can be easily used in the resonant mode as well as in the nonresonant vertical oscillation (NRVO) mode in which the risk of object destruction is greatly reduced. Recall that these techniques are based on the vertical oscillation of the tip which periodically touches the surface. In the resonant mode, the working frequency is a resonant mode of the cantilever whereas in the NRVO mode the tip oscillates with a larger amplitude, out of this frequency range (several nanometers). It has been shown that such a procedure limits the destructive contact (i) because there is no longer rubbing during scanning and (ii) the viscosity of the interaction can dramatically increase because of the velocity of the tip displacement. The analysis of very soft materials (e.g., living cells) is now possible. The contact technique has been mainly used in total internal reflection configuration for which Fig. 5b shows

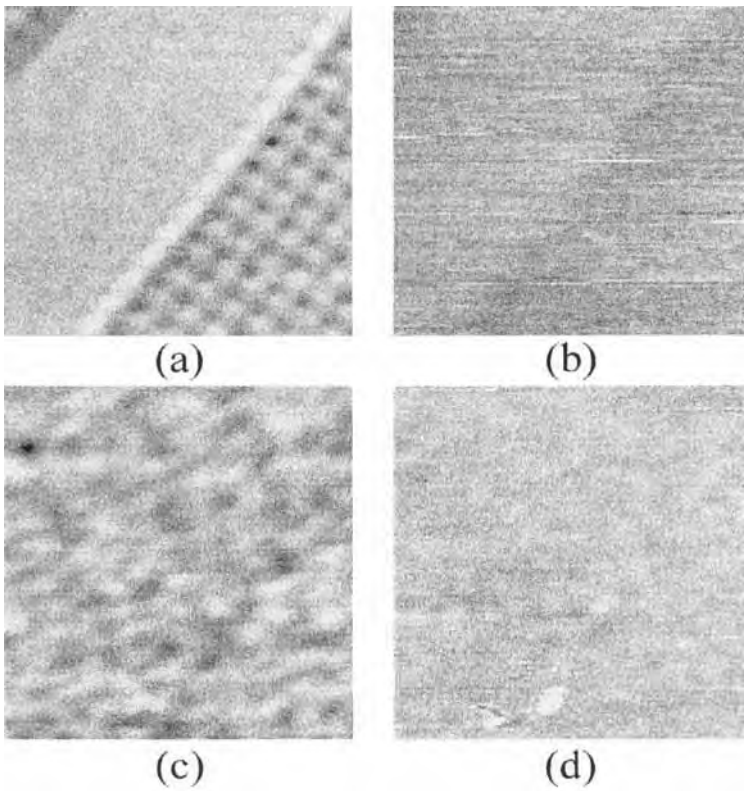


Figure 8. Images of very flat dielectric surfaces exhibiting refractive index variations. (a) Near field optical image of a latent imprint recorded in a photoresist used in lithography (scanned area 6 μm × 6 μm). The so-created index variation is about 10^{-2} and the topography is a few nanometers. The associated shear force image (b) shows no topographic variations. (c) Good contrast optical image of an ion irradiated polymer. (d) Topography variation visible in the shear force image is about 10 nm.

the basic configuration. This technique is directly combined with a force detection by a deflection measurement (dashed lines) or by an interferometric method (full lines).

The total internal reflection microscope has been improved either by using supertips (addition of a protrusion of carbonaceous material to the summit of the silicon nitride pyramid [49]) or by metallizing the tip, leaving its apex free of metal [51]. Some interesting alternatives use metallic tips working in the perturbation mode (Fig. 9a). Good resolved images have been obtained either in the contact mode or better still in vertical oscillation mode [30].

5.5.5 Distance Control by Measurement of the Electron Tunneling Current Between Tip and Sample

In one of the first near field microscope experiments, the distance between the tip and the sample was estimated by measuring the tunneling current [1]. A few attempts have been made in this way and highly resolved near field images have been obtained by combining a nanoaperture microscope and an STM control [52, 53]. Note that the use of purely metallic sharp tips working in the perturbation mode have been used in an STOM configuration

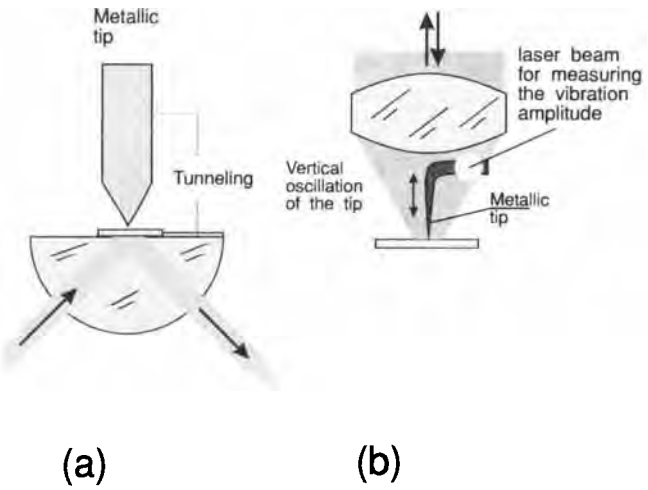


Figure 9. Use of opaque metallic sharp tips. (a) The configuration is an inverted STOM, in which the tip modifies the frustration of the total reflection. The resulting field variation is measured in the reflected beam. The tip can be used as an electron tunneling probe to control the tip–sample distance. (b) Reflection configuration in which the perturbation is collected by means of a high aperture microscope objective. NRVO mode can be associated to prevent tip damages.

(Fig. 9b). Probably the best resolved images have been made that way [28].

5.6 Near Field Microscopy Applications

Although near field microscopy is still a developing technique, its fields of application is now beginning to take shape. The main area of application seems to deal with spatially resolved spectroscopy. Biologists are probably the first users of this local spectroscopy. Besides biology, other applications can be envisaged, for example, in the field of local photoemission. Figure 10 is an illustration of the local analysis of the photoemission in a Schottky barrier.

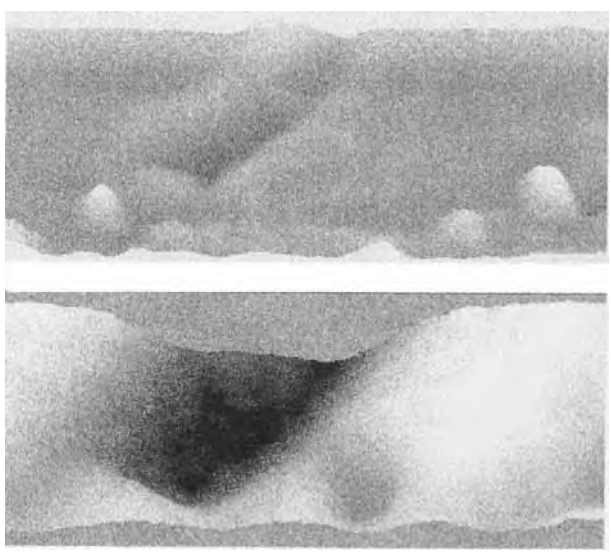


Figure 10. Use of a reflection near field microscope to map the internal local photoemission at a Pt–GaP interface (Schottky barrier). In this experiment the tip locally illuminates the interface and the induced current is simultaneously recorded with the optical and the topographic image. Upper image: the topography measured by the shear force technique (scanned area 14 μm × 5 μm). Lower image: local photocurrent.

Finally, near field microscopy can provide local information about the physical properties of matter on a scale which was inaccessible prior to 1980. Certainly, the increasing knowledge of confined fields is a direct consequence of the development of near field microscopy.

5.7 References

[1] D. W. Pohl, W. Denk, M. Lanz, *Appl. Phys. Lett.* **1984,** *4*, 651.
[2] E. H. Synge, *Phil. Mag.* **1928,** *6*, 356.
[3] J. A. O'Keefe, *J. Opt. Soc. Am.* **1956,** *46*, 359.
[4] E. A. Ash, G. Nicholls, *Nature* **1972,** *237*, 510.
[5] J. Goodman (Ed.) *Introduction to Fourier Optics*, McGraw-Hill, New York **1968.**
[6] A. Blanc-Lapierre, M. Perrot, G. Peri, *C.R. Acad. Sc. Paris* **1953,** *256A*, 1540.
[7] A. W. Lohmann, D. P. Paris, *Appl. Optics* **1964,** *3*, 1037.
[8] M. Pluta (Ed.) *Advanced Light Microscopy*, Vol. 2, Elsevier, Amsterdam **1989**, pp. 357–369.
[9] C. J. Chen, *Ultramicroscopy* **1991,** *42–44*, 1653.
[10] K. Wickramasinghe, C. C. Williams, IBM Patent *344621*, **1989.**
[11] G. A. Massey, *Appl. Optics* **1984,** *23*, 658.
[12] U. C. Fischer, H. P. Zingsheim, *Appl. Phys. Lett.* **1982,** *40*, 195.
[13] A. Lewis, M. Isaacson, A. Harootunian, A. Muray, *Ultramicroscopy* **1984,** *13*, 227.
[14] E. Betzig, A. Lewis, A. Harootunian, M. Isaacson, E. Kratschmer, *Biophys. J.* **1986,** *49*, 269.
[15] D. Courjon, K. Sarayeddine, M. Spajer, *Optics Commun.* **1989,** *71*, 23.
[16] D. Courjon, J. M. Vigoureux, M. Spajer, K. Sarayeddine, S. Leblanc, *Appl. Optics* **1990,** *29*, 3734.
[17] R. C. Reddick, R. J. Warmack, T. L. Ferrell, *Phys. Rev. B* **1989,** *39*, 767.
[18] F. de Fornel, J. P. Goudonnet, L. Salomon, E. Lesnieswka, *Proc. SPIE* **1989,** *1139*, 77.
[19] N. F. Van Hulst, F. B. Segerink, B. Bölger, *Optics Commun.* **1992,** *87*, 212.
[20] S. I. Bozhevolnyi, M. Xiao, O. Keller, *Appl. Optics* **1994,** *33*, 876.
[21] H. Bielefeldt, J. Mertz, I. Hörsch, J. Mlynek, O. Marti, *Spatially Resolved Spectroscopy of a Photoluminescent Layer by Scanning Near Field Optical Microscopy*. Proceedings EQEC'93 (Firenze, Sept. 1993), Technical Digest (Eds.: P. de Natale, R. Meucci, S. Pelli), **1993.**
[22] D. Courjon, M. Spajer, A. Jalocha, S. Leblanc, *Near Field Microscopy and Optical Tunneling Detection*, AIP Conference Proceedings, *241*, *Scanned Probe Microscopy*, Ed.: K. Wickramasinghe, Santa Barbara, CA, **1991.**
[23] D. Van Labeke, D. Barchiesi, *J. Opt. Soc. Am. A* **1993,** *10*, 2193.
[24] R. Winston, *Sci. Am.* **1991,** 76.
[25] D. Courjon, C. Bainier, F. Baida, D. van Labeke, D. Barchiesi, *Near Field Instrumentation*, Workshop NATO, Reichenau Germany **1994**, *Proceedings*, in press.
[26] A. Roberts, *J. Appl. Phys.* **1991,** *70*, 4045.
[27] L. Novotny, D. W. Pohl, P. Regli, *J. Opt. Soc. Am. A* **1994,** *11*, 1768.
[28] J. D. Pedarnig, M. Specht, W. M. Heckl, T. W. Hänsch, *Scanning plasmon near field microscope*. Proceeding NFO NATO, series 242 (Eds.: D. W. Pohl, D. Courjon), **1993,** pp. 273–280.
[29] Y. Inouye, S. Kawata, *Opt. Lett.* **1994,** *19*, 159.
[30] P. Gleyzes, A. C. Boccara, R. Bachelot, *Near Field Optical Microscopy using a Metallic Vibrating Tip*. Proceedings NFO2 (Raleigh Oct. 1993), *Ultramicroscopy* **1995.**
[31] a) F. Zenhausern, M. P. O'Boyle, H. K. Wickramasinghe, *Appl. Phys. Lett.* **1994,** *65*, 1623, b) F. Zenhausern, Y. Martin and H. K. Wickramasinghe, *Science* **1995,** *269*, 1083.
[32] W. E. Moerner, T. Plakhotnik, T. Irngartinger, U. P. Wild, *Phys. Rev. Lett.* **1994,** *73*, 2764.
[33] E. Betzig, R. J. Chichester, *Science* **1993,** *262*, 1422.
[34] E. Betzig, A. Harootunian, M. Isaacson, E. Kratschmer, *Appl. Phys. Lett.* **1987,** *51*, 2088.
[35] U. Ch. Fischer, U. Dürig, D. W. Pohl, *Near-Field Optical Scanning Microscopy and Enhanced Spectroscopy with Submicron Apertures*, Scanning Microscopy International, Chicago (AMF O'Hare), **1987,** pp. 47–52.
[36] U. Ch. Fischer, D. W. Pohl, *Phys. Rev. Lett.* **1989,** *62*, 458.
[37] M. Spajer, D. Courjon, K. Sarrayeddine, A. Jalocha, J. M. Vigoureux, *J. Phys.* **1991,** *3*, 1.
[38] C. Girard, M. Spajer, *Appl. Optics* **1990,** *29*, 3726.
[39] R. D. Grober, T. D. Harris, J. K. Trautman, E. Betzig, *Rev. Sci. Instrum.* **1994,** *65*, 626.
[40] B. Hecht, H. Heinzelmann, D. W. Pohl, *Combined Aperture SNOM/PSTM: the Best of Both Worlds?* Proceedings NFO2 (Raleigh, Oct. 1993), *Ultramicroscopy* **1995.**
[41] O. Marti, H. Bielefeldt, B. Hecht, S. Herminghaus, P. Leiderer, J. Mlynek, *Optics. Commun.* **1993,** *96*, 225.

[42] P. M. Adam, L. Salomon, F. de Fornel, J. P. Goudonnet, *Determination of the Spatial Extension of the Surface Plasmon Field in the Near Field Region of a Silver Film with a Photon Scanning Tunneling Microscope.* Proceedings NFO2 (Raleigh, Oct. 1993), *Ultramicroscopy* **1995**.

[43] D. Courjon, C. Bainier, M. Spajer, *J. Vac. Sci. Technol. B* **1992,** *10*, 2436.

[44] E. Betzig, P. L. Finn, J. S. Weiner, *Appl. Phys. Lett.* **1992,** *60*, 2484.

[45] R. Toledo-Crow, P. C. Yang, Y. Chen, M. Vaez-Iravani, *Appl. Phys. Lett.* **1992,** *60*, 2957.

[46] H. Bielefeldt, I. Hörsch, G. Krausch, M. Lux-Steiner, J. Mlynek, O. Marti, *Appl. Phys. A* **1994,** *59*, 103.

[47] O. Bergossi, M. Spajer, *Proc. SPIE* **1994,** *2341*, 239.

[48] J.-K. Leong, C. C. Williams, *Appl. Phys. Lett.* **1995,** *66*, 1432.

[49] M. H. P. Moers, R. G. Tack, O. F. J. Noordman, F. B. Segerink, N. F. Van Hulst, B. Bölger, *Combined Photon Scanning Tunneling Microscope and Atomic Force Microscope Using Silicon Nitride Probes.* Proceedings NATO, series 242, Eds.: D. Pohl, D. Courjon, Kluwer Acad. **1993,** pp. 79–86.

[50] F. Baida, D. Courjon, G. Tribillon, *Combination of a Fiber and a Silicon Nitride Tip as a Bifunctional Detector: First Results and Perspectives.* Proceedings NFO, (Oct. 1992), NATO series, Eds.: D. W. Pohl, D. Courjon, series *242*, **1993,** pp. 71–78.

[51] F. Baida, *Microscopie Hybride: Association d'un Microscope Optique en Champ Proche et d'un Microscope à Force Atomique. Principe et Réalisation.* Thèse d'Université, *434*, Besançon, France **1995**.

[52] K. Lieberman, A. Lewis, *Appl. Phys. Lett.* **1993,** *62*, 1335.

[53] M. Garcia-Parajo, E. Cambril, Y. Chen, *Appl. Phys. Lett.* **1994,** *65*, 1498.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

6 Infrared Microscopy

6.1 Infrared Spectroscopy

6.1.1 Molecular Vibration Spectroscopy

For organic chemists, spectroscopy is the most widely used physical method for investigating molecular structures. A molecule interacts with electromagnetic radiation to absorb energy which corresponds to transitions between fixed energy levels for the molecular species [1]. The difference in energy between two energy levels is proportional to the frequency of the light absorbed and is expressed as $E = h\nu$ where h is Planck's constant (6.624×10^{-34} J s) and ν is the frequency of the light in hertz. The frequency of the light is related to the wavelength λ by the expression $\nu = c/\lambda$, where c is the velocity of the light ($c = 2.998 \times 10^{8}$ m s^{-1}).

A molecule that has absorbed energy is said to be in an excited state, above its ground state. It returns to the ground state by loss of energy, usually by emitting heat or, less frequently, light. Molecules have several different kinds of energy levels (Fig. 1): first, because a molecule rotates in space as a rigid body, the various types of rotational motion correspond to rotation levels; second, because the atoms within a molecule are in constant motion relative to each other, there are vibrational levels; and finally, just as atoms have energy levels among which electronic transitions take place, so do electrons which are involved in chemical bonds.

The gaps between electronic levels are much larger than those between vibrational levels, and energy differences between rotational levels are the smallest. The differences in electronic, vibrational, or rotational energy levels correspond to energy in different parts of the electromagnetic spectrum. In the following account, we will focus on the infrared region ($2.5\,\mu m < \lambda < 25\,\mu m$) to investigate how molecular interactions with radiation in that energy range (about 20 kJ mol^{-1}) give us information about molecular structure.

The atoms within a molecule are constantly in motion, distorting the chemical bonds. These motions are called molecular vibrations [2]. One type of vibration, the 'stretching vibration', causes changes in bond length. Other vibrations cause changes in bond angles and are called 'bending vibrations'. The different vibrational modes are described by a number of energy levels and the spacing between these levels corresponds to the energy of the infrared photons, given in units of wavenumbers ($\bar{\nu} = \lambda^{-1}$ from 4000 to 400 cm^{-1}).

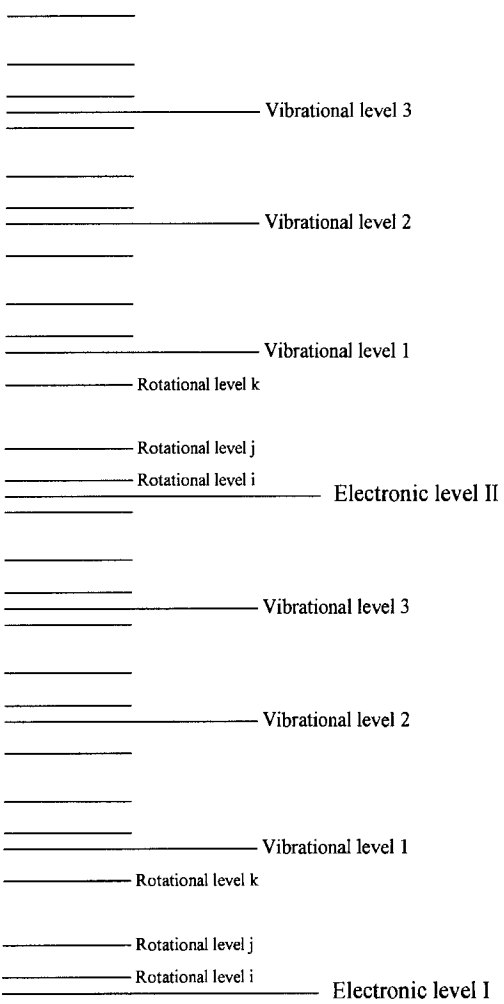


Figure 1. Molecular energy level diagram.

When radiation having the same energy as the difference in energy between two energy levels strikes a molecule, the radiation is absorbed and the amplitude of the molecular vibration increases. This energy, which is not enough to break chemical bonds or cause chemical reactions, is ultimately returned to the environment as heat energy.

Since there are $3N - 6$ vibration modes for a molecule with N atoms ($3N - 5$ for a linear molecule), an infrared spectrum usually has many absorption bands which correspond to a large number of possible molecular vibrations, even if the molecule is simple. The spectrum is useful for a chemist because the different functional groups absorb at different frequencies which are related to the types of bonding present in them [3]. The major frequencies typical of functional groups (Table 1) usually lie between 4000 and 1400 cm^{-1}. The portion of the spectrum under 1400 cm^{-1} is called the fingerprint region because the bands are characteristic for a given molecule. It is also more difficult to make specific assignments of bands in that region as they are more dependent upon the structure of the molecule as a whole. This is immensely useful for unmistakably identifying a compound. For two samples, the presence of all bands with the same relative intensity, in the fingerprint as well as in the higher frequency region of the spectrum, is considered to be proof of the identity of the two.

Not all bands in the infrared spectrum have the same intensity, and this too is useful in identifying functional groups. During a vibration that corresponds to a change in the dipole moment, the molecule absorbs radiation strongly. For example, this is the case for the stretching of a carbonyl group. On the other hand a vibration in which there is a small or no change in the dipole moment is observed as a weak absorption, such as the stretching of a carbon–carbon bond.

6.1.2 Instrumentation

The development of infrared spectrometers was achieved during the first part

Table 1. Characteristic infrared absorptions of organic compounds

Spectral region [cm^{-1}]	Vibrational mode	Wavenumber [cm^{-1}]	Intensity*
above 2700	O–H stretching	3650–3590	m
	O–H stretching (hydrogen bonded)	3600–3200	s, broad
	N–H stretching	3500–3300	m
	C–H alkyne stretching	3300	s
	C–H alkene stretching	3080–3020	m
	C–H aromatic stretching	3100–3000	m
	C–H alkane stretching	2980–2850	s
	C–H aldehyde stretching	2800–2700	m, 2 bands
2500–2000	C≡C stretching	2260–2100	m
	C≡N stretching	2260–2220	m
1800–1600	C=O ester stretching	1750–1735	s
	C=O acid stretching	1725–1700	s
	C=O aldehyde stretching	1740–1720	s
	C=O ketone stretching	1725–1705	s
	C=O aryl ketone stretching	1700–1680	s
	C=O amide stretching (amide I mode)	1690–1650	s
	C=C alkene	1680–1620	m
1600–1300	C=C aromatic	1600–1500	m
	N–H amine bending	1600–1550	w
	N–H amide bending (amide II mode)	1550–1500	m
	C–H alkane bending	1470–1350	s
1300–1000	C–N amine stretching	1300–1200	s
	C–O ether, ester stretching	1300–1100	s
	C–O alcohol stretching coupled with O–H bending	1300–1000	m
below 1000	C–H alkene bending	1000–800	s
	C–H aromatic bending	850–700	s

* s = strong, m = medium, w = weak

of this century and World War II brought greater need for analytical instruments. Significant progress in electronics enabled the construction of IR gratings or prism spectrometers with rapid response detectors. So it became possible to record IR spectra in a few minutes. However, conventional grating spectrometers suffer a number of drawbacks: (i) the sensitivity of conventional grating spectrometers is insufficient to record good quality spectra of low concentrations, small or strongly absorbing samples; (ii) measurements are slow, and only specially designed spectrometers can measure the complete mid-infrared spectrum in less than a few minutes with good resolution and signal-to-noise ratio; and (iii) the fundamental drawback of a prism or grating spectrometer is found in the monochromator which limits the frequency range of the radiation reaching the detector to one resolution width. As the resolution is increased less energy reaches the detector [4].

Fellgett [5] was interested in astronomical observations of extremely low radiation intensity. He investigated methods of multiplexing the spectral information and realized that the use of a

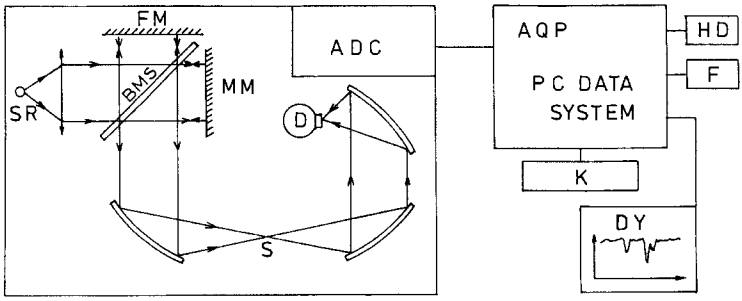


Figure 2. Schematic representation of the FTIR spectrometer. SR: source, BMS: beam-splitter, FM: fixed mirror, MM: mobile mirror, S: sample, D: detector, ADC: analogic/digital converter, AQP: acquisition processing, K: keyboard, HD: hard disk, F: floppy, DY: display.

Michelson interferometer could yield a great reduction during the time taken to measure the spectra of weak sources. The reduction in measurement time, which results from measuring all of the radiation during all of the measurement, is known as 'Fellgett's advantage'. Another advantage in the use of a Michelson interferometer for spectroscopy investigation concerns the energy throughput. The increased signal-to-noise ratio resulting from the increased signal at the detector is called 'Jacquinot's advantage'. So spectrometers using a Michelson interferometer instead of a grating monochromator were developed. But initially Fourier transform spectroscopy was not attractive to chemists because of the inconvenience of the computation of the spectrum from the interferogram [4].

Fourier transform infrared spectroscopy (FTIR spectroscopy) experienced a great boom when more powerful and faster computers were developed as a result of advances in solid-state device technology and integrated circuits. From 1968 commercial FTIR spectrometers allowed the recording of infrared spectra in a few seconds with good resolution and good signal-to-noise ratio. Thus, for analytical chemists and chemical spectroscopists the FTIR spectrometer has become a routine, most widely used tool for the identification of organic and inorganic compounds.

The optics and PC data system of a modern FTIR spectrometer are schematized in Fig. 2. The interferogram measured by the detector is amplified and converted by an analogic–digital converter. The PC system can accumulate this signal and converts the interferogram to a spectrum with the use of a fast Fourier transform calculation. With this arrangement it is possible to obtain good quality infrared spectra (sensitivity, resolution).

An article by Barber, Cole and Thomas, written in 1949, first reported the linkage of a microscope with an IR spectrometer [6]. These researchers reported their ability to record spectra of crystals or fibers with a diameter of 20 to 50 μm over the full range of the mid-IR spectrum. In 1953 the first commercial infrared microscope attachment for IR spectrometry was produced by Perkin–Elmer, although it was not a commercial success. The principal reason was due to its poor sensitivity. At the 1983 Pittsburgh conference, an infrared microscope device that was specifically designed for Fourier transform spectrometer was introduced by BioRad Digilab Division. This technique then became widely used in many fields.

6.2 Infrared Microspectrometry

6.2.1 Principles

The idea behind FTIR microspectrometry is to use visible light to view a magnified image of the sample and then to select microscopic areas of the sample for infrared spectroscopy analysis. When heterogeneous samples are analyzed by FTIR microspectrometry the goals are to determine qualitatively what constituents are present and to assay quantitatively the amount of each component according to the vibrational molecular spectra.

6.2.1.1 Optical Design

The general layout of the infrared microscope consists of transfer optics to guide the infrared radiation from the interferometer through the microscope [7]. The optical scheme of the A590 Bruker FTIR microscope (Bruker Analytische Messtechnik Gmbh, Karlsruhe, Germany) is shown in Fig. 3. The sample is first observed by transmission or reflection in visible light. The area of interest is brought to the center of the optical field and delineated with high contrast apertures in the image planes of the sample. The visible light is then substituted by the infrared beam which is made colinear with the aid of a series of small pinholes. The transmitted or reflected infrared radiations are finally detected by the IR detector to record the absorption spectrum of the analyzed area.

The main difference with regard to the traditional microscopes, that employ as optical devices combinations of many

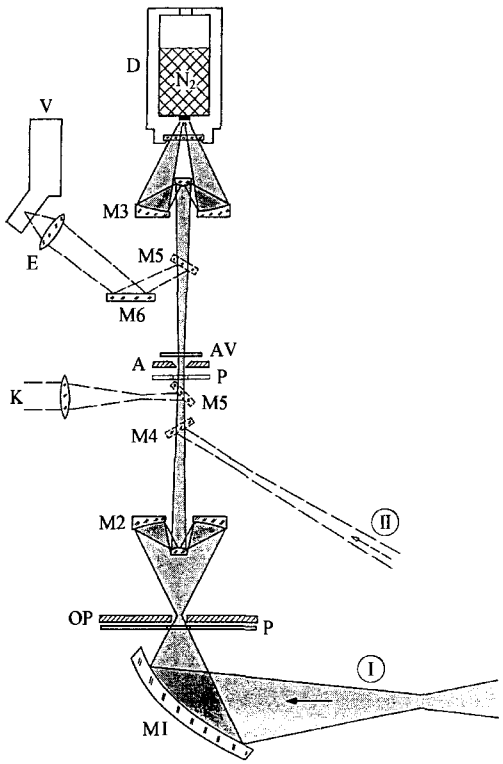


Figure 3. General layout of an FTIR microscope. I: IR transmittance beam, II: IR reflectance beam, M1: condenser mirror, M2, M3: Cassegrain objectives, M4: movable semitransparent mirror, M5: movable mirror, M6: mirror, OP: object plane, A: aperture, E: eyepiece, D: detector liquid N_2 cooled.

lenses, is the limitation of lens material for use in both the visible and the mid-infrared regions. Therefore infrared microscopes use all-reflective objectives. Reflective surfaces have been used for many years in telescopy and they present no major problems for the optical designer. The on-axis mirror systems, that give the best solution to the obscuration problem, are known as the Schwarzschild configuration.

At this stage, it is important to remember the advantages which are offered by FTIR technology in comparison with dispersive instruments. The multiplex advantage or the high energy throughput

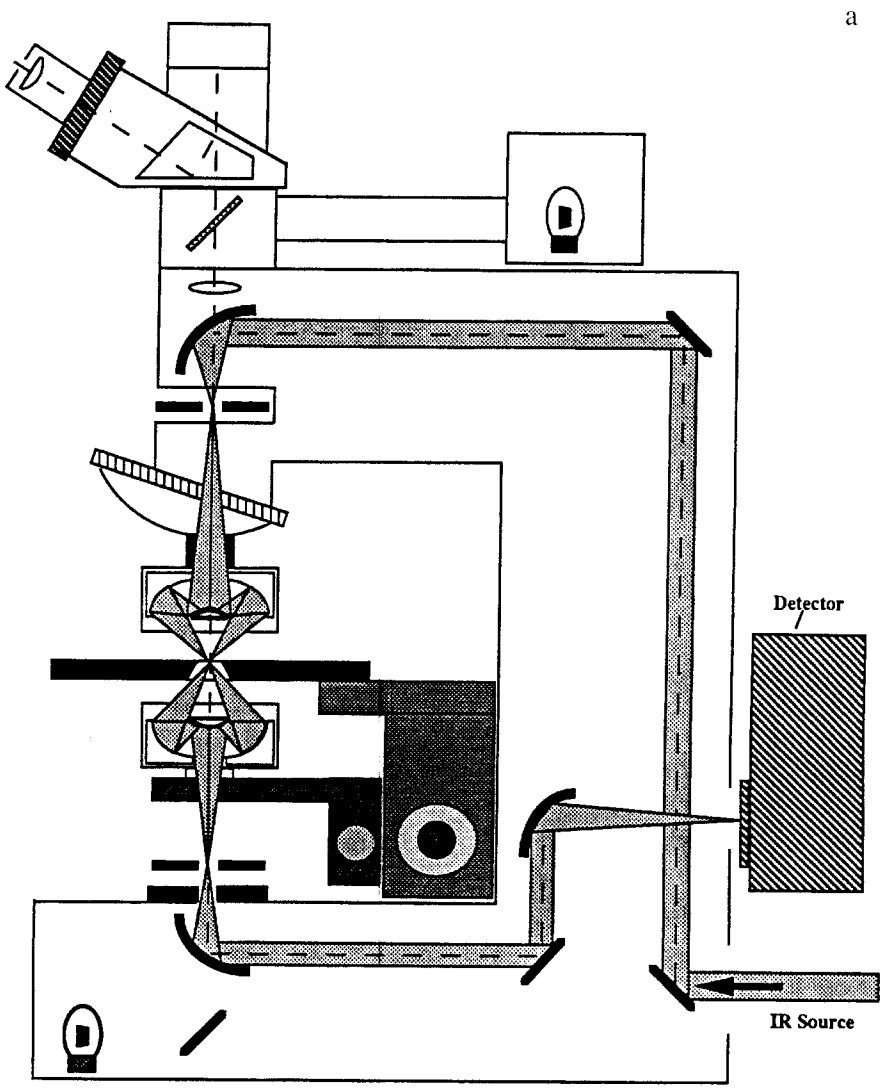


Figure 4. Operational modes of an FTIR microscope. (a) Transmission mode; (b) Reflection mode.

constitute appreciable conditions for the spectral analysis of the weak IR signal collected through a very small sample. The technology employed in the infrared detector is another particularity of the infrared microscope design: namely, it needs a very sensitive detector whose elemental dimensions must be in good agreement with the beam size. Typically, we use a $0.25\,mm^2$ narrow band MCT detector whose performance will be discussed along with signal-to-noise ratio considerations.

6.2.1.2 Operational Modes

The transmission and reflection modes [8] are illustrated in Fig. 4 which reproduces

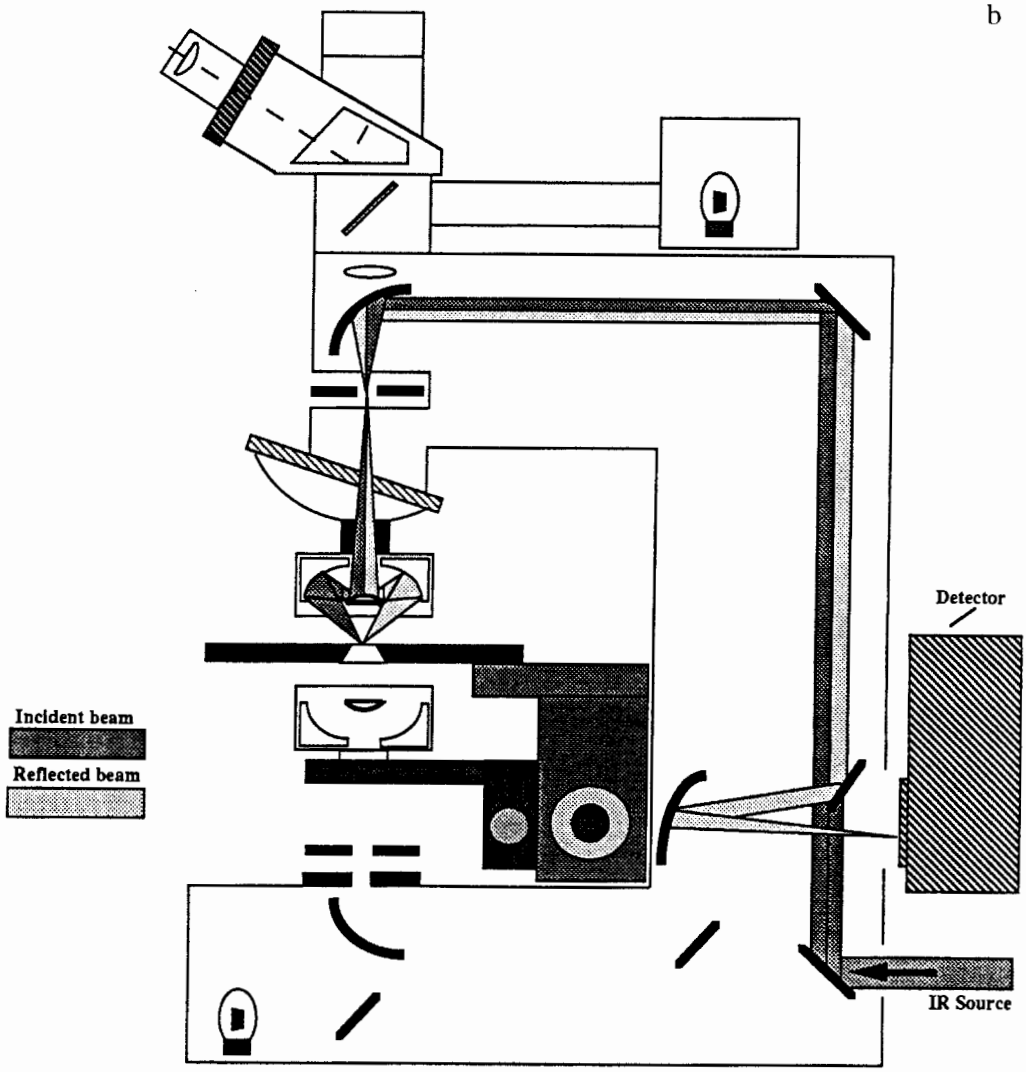


Figure 4. Continued

the schematic diagram of the IR-Plan Spectra-Tech microscope (Spectra-Tech, Inc., Stamford, CT, USA).

(a) *Transmission*

IR radiation from the interferometer passes through the upper aperture and is focused by the reflective objective onto the sample plane. The radiation is collected by the condenser, passes through the lower aperture and reaches the detector. Along this light path, different optical phenomena may appear.

The diffraction modifies the light direction when passing through a very small aperture. Since this physical phenomena depends on the wavelength of

the radiation, it is critical in the infrared range and must be taken into account to determine the spatial resolution limit.

The refraction effect may be observed when IR transmitting windows are used to support the sample. This effect induces a contribution of the neighborhood of the analyzed area to the absorption spectrum. Precise adjustment of the objective and condenser is necessary to preserve the spatial resolution.

Internal reflections may occur in a flattened sample. This appears as an interference fringe effect since the wavelength is often on the order of the thickness of the thin film. The probability of internal reflection may be reduced by mounting the sample between two windows of which the refractive index is closer to the sample's refractive index than is air.

Scattering due to inhomogeneity causes a loss of energy. This effect is often greater at higher wave numbers and hence leads to a slope in the spectrum baseline. This may be corrected by software.

(b) *Reflection*

Since many samples, such as thick samples, thin films on metallic substrates, surface contaminants etc., cannot be studied easily by the transmission mode, infrared reflection spectroscopy allows these measurements with reduced sample preparation. However, interpretation of the reflectance spectra is more difficult since there are several kinds of reflected signals which may simultaneously contribute to the recorded spectra.

Specular reflectance is front-surface reflection. The beam which does not penetrate the material is reflected according to the Fresnel laws. Specular-reflectance spectra are often very different from absorption spectra. The bands appear like first-derivative functions. It is a consequence of variations of two separate optical parameters, the refractive index and the absorption coefficient. Before interpretation, the data has to be resolved via the Kramers–Kronig equation to extract the absorption contribution. The position and the intensity of the Kramers–Kronig transformed reflectance spectra are well-matched with transmission data. Hence, identification by comparison to spectral libraries or quantitative applications are possible.

Diffuse reflectance may be considered as a result of multiple reflection, refraction and diffraction processes on solid particles which are randomly oriented. This reflection is delocalized and characteristic of a matt surface. The resulting spectra have an appearance more similar to transmittance spectra than specular-reflectance spectra. The position and the shape of the bands may be directly interpreted. For quantitative applications, diffuse-reflectance spectra need an intensity correction by the Kubelka–Munk equation to obtain a good correlation to the concentrations.

Internal reflection spectroscopy is observed when the infrared radiation is passed through a transmitting crystal with a high refractive index. An evanescent wave extends beyond the surface of the crystal into the sample held in contact with the crystal and interacts with the sample which absorbs energy. An infrared spectrum is observed. This technique, which is termed attenuated total reflectance (ATR) spectroscopy, can be applied to all kinds of solid and liquid samples which allow a good contact between them and the crystal. Taking into account the effective path

length into the sample, which depends on the radiation wavelength, the ATR spectrum is a directly usable measurement of the absorption sample spectrum.

Reflection–absorption occurs when a thin absorbing layer of material is on the surface of a high reflective substrate such as a polished metallic surface. The infrared radiation passes twice through the absorbing film with an intermediary reflection on the substrate. This method, known as 'double transmission', characterizes the contaminants on surfaces and leads to very high signal-to-noise ratio spectra due to the collection efficiency. Nevertheless, at near-normal incidence, the sensitivity of the reflection–absorption interaction for films less than 1 μm thick is very small with regard to transmission measurements whereas the infrared radiation passes through the film twice. This result justifies the measurement of thin film absorption at grazing-incidence, between 60 and 90°, since the electric field strength is enhanced at the surface of the metallic substrate.

6.2.1.3 Imaging Capabilities

Supplementary to the interest of IR microspectrometry for characterization of microscopic material, this technology has been developed to allow images of a material to be obtained based on functional group maps [9]. When a sample is studied by transmission or reflection–absorption, the spectral intensities calculated in absorbance units follow the Beer–Lambert law. For a pure compound, we can assume that the measured absorbance at an absorbed wavelength is proportional to the thickness of the compound. Thus, the representation of the absorbance with respect to the sample dimensions constitutes a map of the local abundance of the absorbing material.

For infrared microscopic imaging analysis, the basic configuration of the FTIR coupled microscope is completed by a computer controlled x, y mapping sample stage. This stage can be moved in a one or two-dimensional direction, by steps down to as small as 10 μm. The infrared spectra are recorded at various desired spatial coordinates. As soon as the spectra have been collected through a given aperture and as a function of the x, y spatial dimensions, the absorbance is calculated at a given frequency corresponding to a particular functional group or on a larger spectral window for a more general chemical approach. The plotted absorbance against x, y coordinates constitutes an IR image of the sample. All the images can be displayed in false color (or grey levels) contour or surface projection modes. The desired form of representation depends on whichever type of analysis required from the data representation is easiest to interpret. In this context, it is important to remember that the spectral absorbance not only depends on the thickness but also on the absorption coefficient of the sample when comparing the IR images recorded at different wavelengths on different chemical compounds.

IR imaging is used in many fields of science including, for instance, polymer science, materials science, biology, medicine, and chemistry. The images presented in Fig. 5 illustrate the analysis of inclusions in polymer laminate. They were measured with a Perkin–Elmer i-series FTIR microscope connected to a PE system 2000 FTIR spectrometer (Perkin–Elmer Corp.,

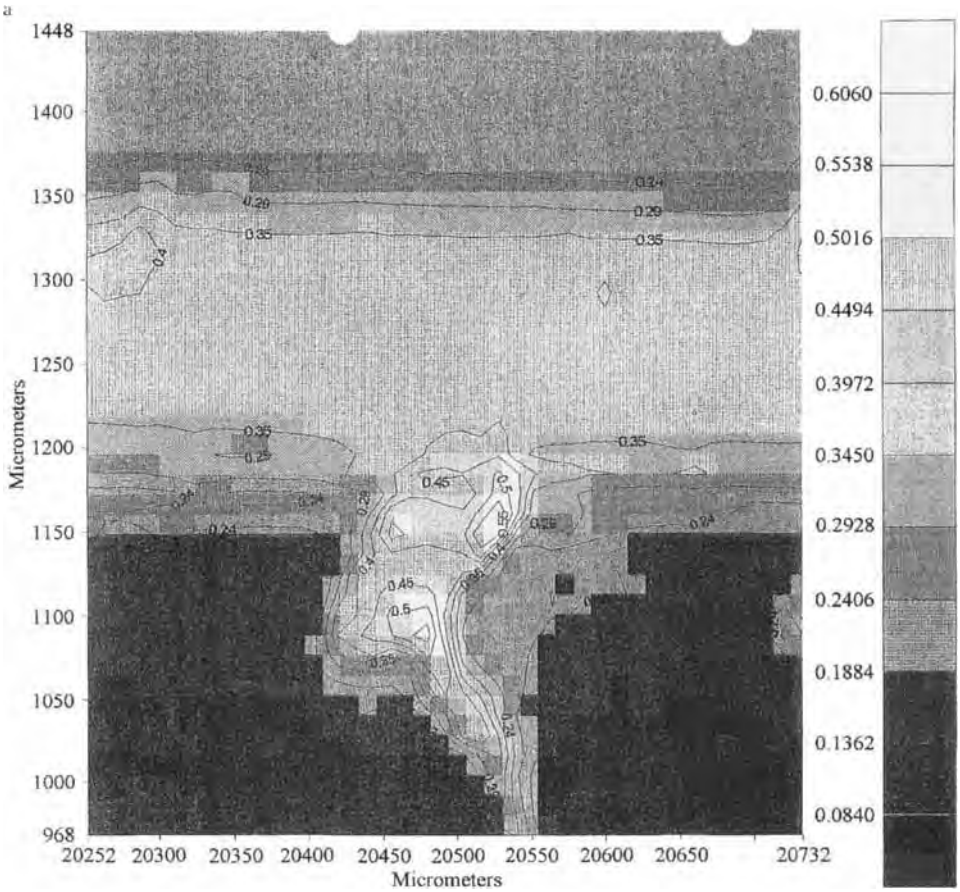


Figure 5. IR imaging application to the analysis of inclusion in polymer laminate. (a) An absorption map.

Norwalk, CT). The sample was prepared as a 100 µm slice using a microtome and the measurements were made in transmission mode with a stage step equal to 12 µm along the x and y directions. The absorbance is calculated on the whole IR domain. So, the two maps in false color and in 3D modes are not specific for any functional group. Nevertheless, the transmission spectrum recorded on the point which coordinates are 20468 and 1148 shows the presence of a carbonyl band and a 1735 cm^{-1} single wave number plot makes the inclusion more evident.

6.2.2 Limitations

6.2.2.1 Spatial Resolution

Spatial resolution is the most important performance criterion to consider when evaluating an FTIR microscope. In a microscope measurement the IR radiation is sent through a tiny opening so that diffraction effects must be considered in contrast with the FTIR spectrometer compartment for macroscopic measurement where the beam is large enough and the diffraction effects are negligible.

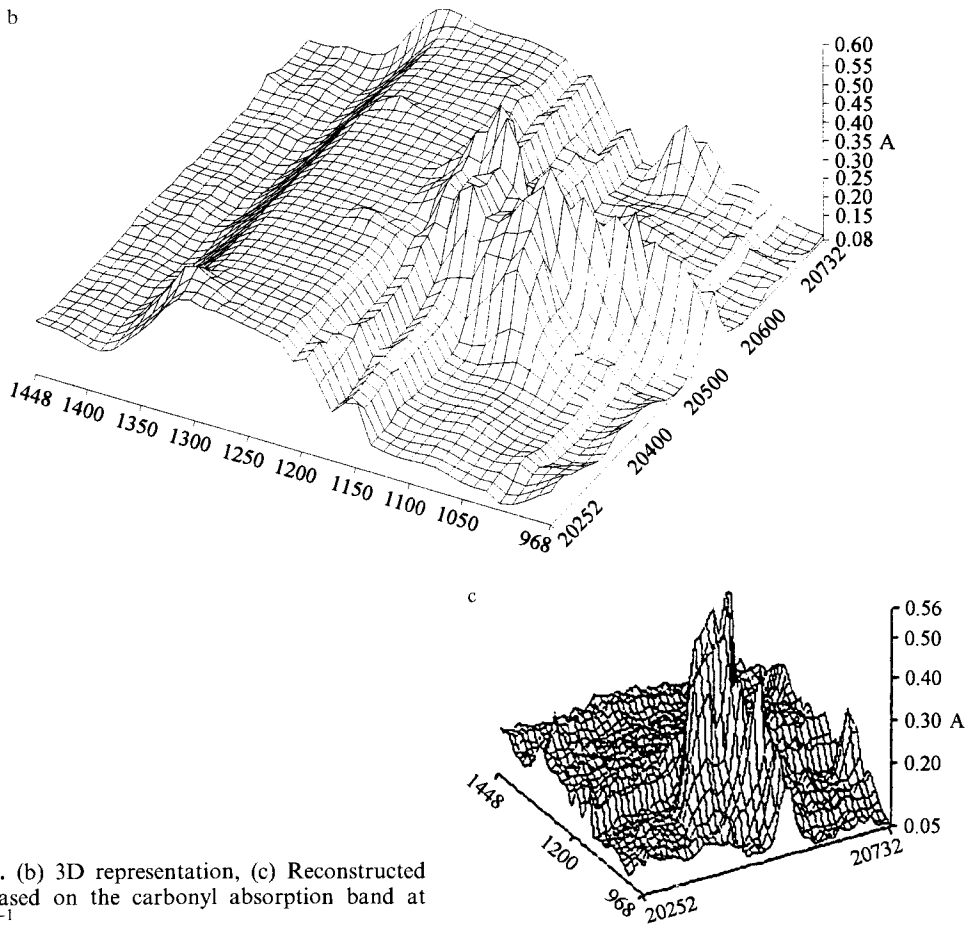


Figure 5. (b) 3D representation, (c) Reconstructed image based on the carbonyl absorption band at 1735 cm^{-1}.

Diffraction occurs when light passes by the edges of opaque bodies or through narrow slits, resulting in light intensity appearing in the geometric shadow of a high contrast edge. Stray light (also called 'spurious energy') is defined as energy that reaches the detector but which does not originate from the geometric area defined by the aperture or the slit used. Diffraction, which varies with frequency, is the major cause of stray light in FTIR microscopy measurements. The effect of stray light on infrared spectral data is to cause erroneously low absorbance values.

A conventional FTIR microscope lacks a targeting aperture between the IR source and the sample. Due to diffraction effect the source can illuminate a sample area greater than that delineated by the aperture so the effective spatial resolution is smaller than expected. To reject a large percentage of the unwanted radiation a second aperture is added in some microscopes (Redundant Aperturing). Double aperturing is used to minimize the diffraction effect. The lower aperture rejects the diffracted radiation, thereby ensuring that the infrared spectrum will contain

information only from the specified area of interest. Thus, with a conventional FTIR microscope with only an aperture between the IR source and the sample, the best spatial resolution that can be achieved is $\geqslant 10\,\mu m$. For better resolution Redundant Aperturing is the solution.

6.2.2.2 Optical Aberrations

Reflecting microscope objectives are being used in applications other than conventional microscopy such as FTIR spectroscopy and FTIR microscopy. These are based on a two mirror design introduced by Schwarzschild [10]. The system, a sort of inverse Cassegrain system, consists of a small primary convex mirror and a large secondary concave mirror. If both mirrors are spherical with a common center, proper spacing results in a system free of spherical aberrations, coma, and astigmatism.

In FTIR microspectrometry the sample acts as an optical element itself and introduces optical aberrations (the sample can be a cylindrical fiber, a solid deposited on a KBr disk or BaF_2 window, a sample pressed between two diamonds and so on). To solve this problem the separation of the two mirrors of the reflecting objective may be adjusted by means of a moving collar and spherical aberrations can be corrected. But chromatic aberration persists and serious chromatic aberration is introduced if the sample support is too thick.

6.2.2.3 Signal-to-Noise Ratio

The signal-to-noise ratio of a spectrum measured with an FTIR spectrometer is defined by the expression [4]:

$$SNR = \frac{u_\nu(T) \times \Theta \times \Delta\nu \times t^{1/2} \times \xi}{NEP}$$

where $u_\nu(T)$ is the spectral energy density for a blackbody, Θ the limiting optical throughput, either at the detector or interferometer, $\Delta\nu$ the resolution, t the time of measurement, ξ the overall system efficiency, and NEP the noise equivalent power of the detector.

For an FTIR microspectrometer the expression of the signal-to-noise ratio can be deduced from the preceding expression [7]:

$$SNR = \frac{u_\nu(T) \times \Theta \times \Delta\nu \times t^{1/2} \times \xi \times \delta}{NEP}$$

where ξ is now the efficiency of the overall system with the microscope, and δ is a new term, the efficiency as influenced by diffraction.

The specimen limits throughput into the FTIR microscope. The throughput at the specimen, Θ, is the product of the sample area and the solid angle subtended by the objective and can be expressed as a function of the sample area and the numerical aperture of the objective.

To compensate the loss of signal-to-noise due to the three terms Θ, ξ, and δ a narrow band MCT detector with an $NEP \approx 5 \times 10^{-13}$ is used in the microscope.

6.2.3 Specific Accessories

6.2.3.1 Sample Preparative Accessories

In FTIR microscopy analyses, sampling is often a quite long and difficult process because in most cases initial sample sizes

are not suitable. For transmittance study the thickness must be chosen to allow the energy reaching the detector to be sufficient. In this way, it is possible to record spectra with the best signal-to-noise ratio without the saturation of the most important bands.

Several preparative accessories are available such as microtome cutting, hydraulic or hand presses, and diamond anvil cell. In particular, a microcompression diamond cell permits the crushing and flattening of hard materials such as minerals or polymers. Two parallel diamond windows that have no facets are used to flatten the sample to an ideal thickness for transmission analyses. This permits multiple samples to be pressed, easily observed and the infrared spectrum recorded.

6.2.3.2 Infrared Objectives

In addition to the transmission and external reflection objectives some manufacturers have developed specific infrared objectives suitable for ATR and grazing incidence analyses.

(a) *Grazing Incidence Objective* [11]: This provides infrared radiation at grazing incidence (65–85°) for the analysis of ultrathin (submicrometer) coatings on metallic substrates. The analysis is performed on a microscopic area of 50 μm × 50 μm. For the measurement in plane (*p*) and perpendicular (*s*) components of reflectivity an infrared polarizer may be added.

(b) *ATR Objective* [12]: The ATR objective permits the direct in situ analysis of highly absorbing materials. The standard objective uses a high refractive index zinc selenide crystal. The sample is positioned in contact with the ATR crystal. Germanium has a high refractive index causing a shallow depth of penetration compared with ZnSe. Other crystals are used (e.g., diamond) if the sample is on a hard substrate.

6.3 FTIR Microsampling

The most often used FTIR microsampling methods will now be illustrated through the description of several significant applications of FTIR microscopy. Some applications have been achieved in our laboratory and others are extracted from the literature in which there has been an abundance of applications over the last ten years. These examples deal with fundamental research or industrial control of materials and they are classified with regard to the operational modes previously defined in the text.

6.3.1 Transmission Studies

6.3.1.1 Characterization of Polymers and Contaminants

For thin film polymer samples the spectrum may be recorded by direct transmission through the film placed on the microscope sample stage. This method is relatively easy if the thickness of the film is less than about twenty micrometers. Since this is infrequent, cutting methods traditionally used in microscopy are used to prepare the infrared samples. As these methods are well known, this approach

was reported very early on to reveal, for instance, disturbances in film profiles. In 1985, Curry, Whitehouse, and Chalmers [13] used a 50 μm pinhole aperture to isolate areas of a thin section of a film from a blend of polypropylene and an ethylene–propylene rubber and were able to show by difference spectroscopy that poor dispersion of the rubber phase was responsible for an unsatisfactory film profile. In this case, the sample often has a small size and it must be placed on a transparent IR window. If the sample is absolutely dry, such as with polymer cuts, we can use polished KBr crystals. Multilayered polymer laminates have been the subject of recent FTIR microsampling studies [14], particularly where thin internal layers are inaccessible to measurement. Pouches used to dispense liquid products are constituted by such a construction which represents a flexible barrier structure LDPE–Al–PET–EVA–PET–Al–LDPE (LDPE: low density polyethylene, PET: poly(ethylene terephtalate), EVA: ethylene–vinyl acetate copolymer, Al: aluminum). Its gas barrier properties depend on the presence of a metallized PET layer either side of a central heat-sealable layer. Infrared spectra of a thin section, microtomed from the edge of the pouch, easily reveal the layers LDPE and EVA since their thickness is about 30 μm. For the 12 μm thick PET layers, the spectra contain contributions of adjacent layers since the thickness is in the region of the spatial resolution limit. The spectrum of this layer after subtraction of the EVA and LDPE contributions can be identified as PET, despite the spectral distortions due to the mathematical operation. Classical preparations of microscopic sections, such as embedding in paraffin or resin, can be used but it is essential to exercise care during cutting to avoid spreading out the embedding material on the section. Otherwise, the absorption spectrum would overlap with the embedding material spectrum. The transmission method can be used for fiber investigations either by putting and flattening the fibers between two transparent windows or by cutting cross sections after embedding the fibers into a matrix.

When contaminants of polymers are observed with visible illumination, the IR spectrum can be recorded from the inclusion in a film. Nevertheless it is often reported that the particle, during a routine optical microscopic examination, is isolated with a tungsten needle and transferred onto the IR window. The sample area is then selected optically using a variable aperture before the spectrum is recorded.

With a similar protocol, we have worked on the identification of solid contaminants in mineral water. With a visible microscopic observation, two kinds of particulates were observed in suspension. These were thin chips with an apparent plane surface of about 1 mm^2 and translucent filaments whose length was of the order of 2 mm. The particulates were extracted from the liquid medium with a needle. They were put on a BaF_2 window since this material is not altered by water. After drying, the transmission spectra presented in Fig. 6 were recorded. For the chip analysis, we selected a 36 μm diameter area. The IR absorption (Fig. 6a) is intense due to the thickness of the cutting. A better spectrum was recorded with the filament (Fig. 6b) since it is less thick. Yet, comparison between the two spectra shows that, independently to their different forms, they are constituted by the same

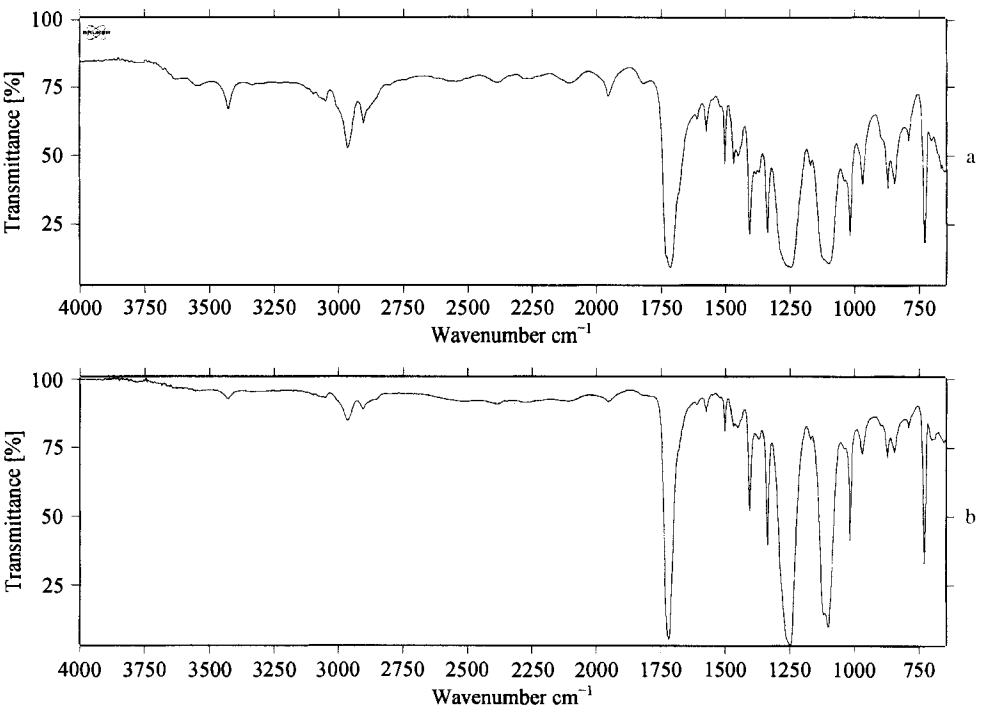


Figure 6. Characterization of solid contaminants in mineral water after removing and drying. (a) Micro-FTIR spectrum of a chip, (b) Micro-FTIR spectrum of a filament.

material. This sample was identified as PET [poly(ethylene terephtalate)] by interpretation of the two spectra with reference to spectral libraries. On the basis of this result, the problem was easily correlated to an error during the packaging process of the mineral water.

6.3.1.2 Direct Structural Identification of Polysaccharides from Red Algae [15]

Because of their gelling ability, agar polysaccharides are widely used in many industrial domains, in the food industry and bacteriology. This physical property is due to their chemical structure. The idealized backbone of agar is made up by an alternating repeating sequence of 3-linked β-D-galactopyranose and 4-linked 3,6 anhydro-α-L-galactopyranose. The neutral form of these polymers is called agarose and has the maximum gelation, but it may be substituted to varying degrees with sulphate esters, *O*-methyl or pyruvate keletal groups which influence the gelation behavior. This kind of polygalactanes comes mainly from the matrix part of the cell wall of red algae.

The reported work has been undertaken to identify agars without any extraction directly in various seaweeds using the infrared microspectrometry method. The main advantage of this method is that the sample consists of a dehydrated algal section only.

Experimentally, the sections come mainly from *Gracilaria verrucosa* collected

in spring. They were obtained from the fresh thallus and next dehydrated with a gradual series of ethyl alcohol and acetone, and dried at a critical point. The final sections have a thickness of about 10 μm. The FTIR spectra were recorded on a Bruker IFS 88 spectrometer coupled with a Bruker microscope. The working mode was in transmission, the resolution was 4 cm^{-1} and 200 scans were collected. The samples (sections) were put on a BaF_2 window which is transparent in the studied wave numbers range (4000–700 cm^{-1}) or studied in a compression cell with diamond windows Spectra-Tech accessory. The analyzed area had a diameter of about 10 μm when we used a ×36 objective with a diaphragm aperture of 0.45.

On a section of *Gracilaria verrucosa* observed under a microscope (Fig. 7a), the cortical area can be easily distinguished from the medullar area: cortical cells are small and the cytoplasm is present in very weak quantities, while the medullar cells become greater towards the center of the section. The latter are mainly constituted by cytoplasm. Under the Bruker microscope, the observed section consists of a dark ring with a clearer inner part. Agar is the main component of cell walls, but it has been demonstrated that the fibrillar part of the cell wall was also constituted by cellulose, mannose and xylose. We suppose that it may be principally located in the cortical area and the greatest amount of these polysaccharides are expected to be identified by FTIR microspectrometry in this part of the algae sections. Thus we have primarily recorded infrared spectra of the different regions of the section. For this purpose different areas, with diameters of about 10 μm, have been selected successively in the external part (the

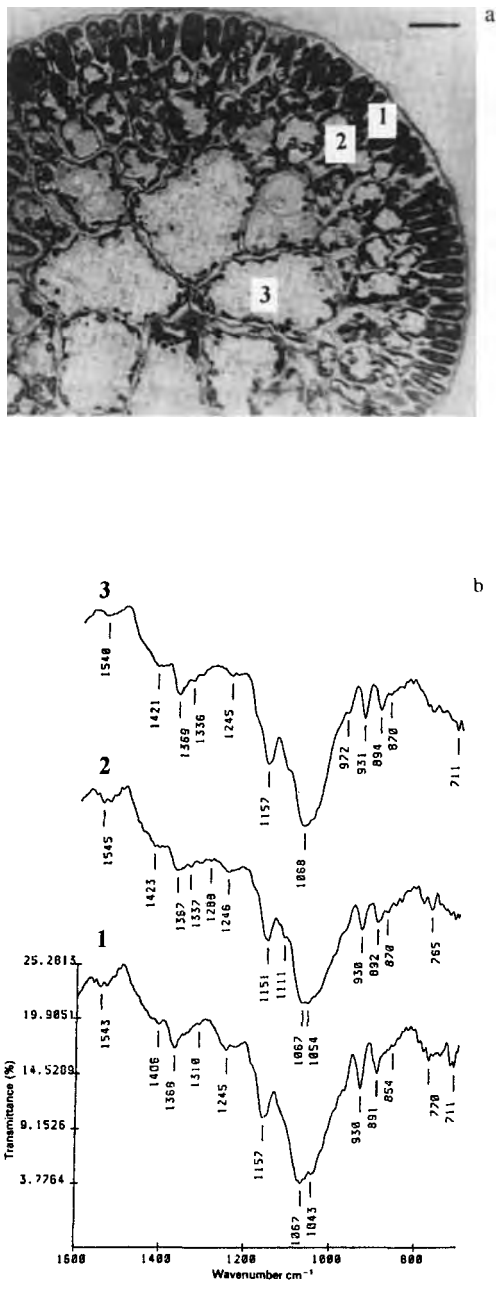


Figure 7. (a) A section of the thallus of *Gracilaria verrucosa*. Scale bar ≡ 2 μm; (b) Micro-FTIR spectra of a cut of *Gracilaria verrucosa*. 1: cortical area, 2: premedullary area, 3: medullary area. (Reproduced from [12] by permission of Springer-Verlag.)

cortical area), the perimedullary and the medullary areas; their spectra were then recorded with the conditions described above (Fig. 7b). The first observation obtained from the different parts of the section is that all the spectra are quite similar and have a general aspect of an agar spectrum. But the precise comparison of the spectra obtained from the cortical area with the agar reference spectrum, demonstrates that the bands present in the spectrum of the agar at about 738 and 716 cm^{-1}, are absent from the medullary area. This observation leads us to conclude that the latter spectrum is essentially the contribution of agar precursors, floridean starch and floridosids rather than the agar absorptions. Another difference between the spectra taken from the three regions is observed. This is the band at about 1250 cm^{-1} due to the substitution by a half ester-sulphate group on C-6, of the 1,4-linked residue constituting the agar polysaccharides. Its presence indicates the presence of agar. On the three spectra, recorded from the cortical, the perimedullary and the medullary areas, we observed that the intensity of the band due to the O=S=O asymmetric stretching mode, became less important from the cortical to the medullary areas. This decrease demonstrates that smaller amounts of agar are contained in the medullary area than in the cortical area. This quantitative difference is investigated for the difference of the cell wall areas which are weaker in the medullary cells because of their big size. The bands at 930 cm^{-1} (C–O–C vibration of the 3,6-anhydrogalactose bridge) and 891 cm^{-1} (skeletal mode due to the β-D-galactose unit in agars) also indicate the presence of agar.

In conclusion about the transmission microsampling methods, we should remember that the sample has to be studied just as it is, without any of the additional compounds usually used for microscopy, such as fixing or coloring agent. The support has to be transparent for the infrared radiation and the presence of water in biological samples must be avoided.

6.3.2 Reflection Studies

We have already seen that specular reflection is the simplest method from the point of view of sample preparation. A block of material may be directly studied by collecting the infrared radiation which is reflected by a polished face. Nevertheless, the Kramers–Kronig transformation needed before specular reflectance spectrum interpretation makes this method very impractical.

In contrast, the study of a thin deposit on a metallic substrate by reflection–absorption is more frequently used. The example presented here concerns some troubleshooting to do with ball-bearings used in high technology equipment. First, it is possible to record the infrared spectrum of the lubricating oil (Fig. 8a) deposited as a thin film on the ball-bearings. After their use under severe conditions deposits in the ball races can be observed. With the FTIR microscope, it is possible to isolate a 50 µm diameter area in the ball race and record the spectrum (Fig. 8b). This spectrum shows the principal absorption bands of the lubricating oil; but, in the 1800–1400 cm^{-1} region, spectral modifications are observed and they may be

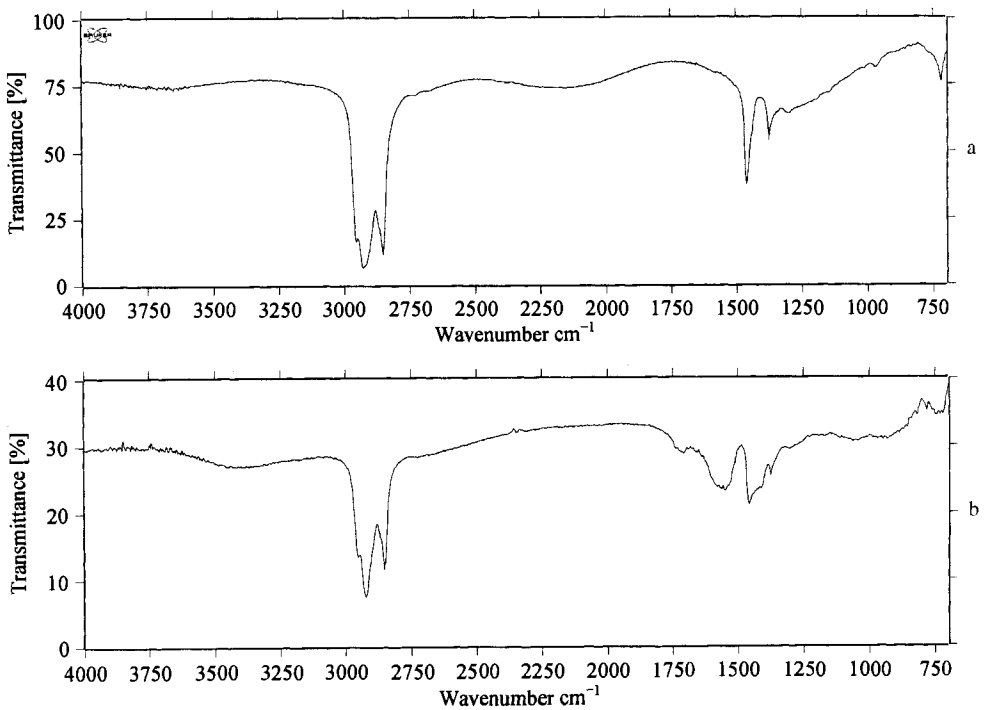


Figure 8. Reflection–absorption IR spectra of (A) the lubricating oil on ball-bearings and (B) the thin deposits observed on ball-races.

assigned to the appearance of degradation products of the oil.

To compensate the principal limitation of the reflection–absorption method, which is a lack of sensitivity when it is measured at normal incidence, a very specific objective described in Sec. 6.2.3 of this Chapter and called grazing-incidence objective can be applied. The constructor reports that surface films and layers as thin as 2 nm can be characterized and mapped. At this scale, the study of monolayer coatings may produce information about the molecular orientation. The technical documentation [11] shows a high quality spectrum of a 200 nm thick silane agent coating on an aluminum substrate. The resolution is 4 cm^{-1} and the number of scans is 250. At about 1200 cm^{-1} the intensity of the Si–C stretching band is 0.15 absorbance unit and consequently the signal-to-noise ratio is very good.

The other specific objective, ATR, represents another new possibility of the microreflectance FTIR spectroscopy. Many forensic applications have been published about identification of fibers, hairs, and automobile paints [16]. The identification of photocopy toner without removing it from the document is a valuable example with regard to time consumption and, by contrast with destructive methods described in previous works where the toner was extracted by a solvent. With the application of microinternal reflectance spectroscopy, spectra may be obtained in a few minutes from photocopies made on different brands of copier.

The total time of analysis for each sample was only two minutes. Spectral differences which indicate different polymer binders were observed. Small bands near 3050, 700, and 670 cm^{-1} indicate a monosubstituted aromatic component, possibly from poly(styrene:butadiene). At 1730 and 1200–1000 cm^{-1} carbonyl and C–O stretching bands are indicative of ester plasticizers and different esters are apparent based on the band positions. The carbon black filler which has a high refractive index change near 1700 cm^{-1} is observed by the dispersion effect as a downward slope in this frequency region.

In conclusion, the FTIR microscope can be used in two ways. It also allows observation with visible light on the microscopic samples under preparation: for example, cells in biological tissues, defects in polymers, contaminants of agrofood media, deposits on metals or semiconductors. Moreover, the microscopic detail may be isolated and a vibrational spectrum recorded by transmission or reflection modes. The interpretation of the absorption frequencies can lead to an identification or at least provide information about the molecular structure of the unknown compound.

6.4 References

[1] J. D. Graybeal, *Molecular Spectroscopy*, McGraw-Hill, New York **1988,** p. 132.

[2] E. B. Wilson, J. C. Decius, P. C. Cross, *Molecular Vibrations*, McGraw-Hill, New York **1955.**

[3] L. J. Bellamy, *The Infrared Spectra of Complex Molecules*, 3rd ed., Chapman & Hall, London **1975.**

[4] P. R. Griffiths, *Chemical Infrared Fourier Transform Spectroscopy*, John Wiley, New York **1975.**

[5] P. Fellgett in *Aspen International Conference on Fourier Spectroscopy 1970* (Eds.: G. A. Vanasse, A. T. Stair, D. J. Baker), AFCRL **1971,** p. 139.

[6] R. Barber, A. R. H. Cole, H. W. Thomas, *Nature* **1949,** *163*, 198.

[7] R. G. Messerschmidt in *Infrared Microspectroscopy: Theory and Applications* (Eds.: R. G. Messerschmidt, M. A. Harthcock), Marcel Dekker, New York **1988,** Chapter 1.

[8] J. Reffner, *Principles of Microscopical IR Spectrometry*, Workshop of the American Academy for Forensic Sciences, San Antonio, TX, USA **1994,** Chapter 5.

[9] M. A. Harthcock, S. C. Atkin in *Infrared Microspectroscopy: Theory and Applications* (Eds.: R. G. Messerschmidt, M. A. Harthcock), Marcel Dekker, New York **1988,** Chapter 2.

[10] W. J. Smith in *The Infrared Handbook* (Eds.: W. L. Wolfe, G. J. Zissis), Office of Naval Research, Washington DC **1978,** p. 20.

[11] Spectra-Tech Product Data Sheet, *Grazing Angle Microscopy*, Spectra-Tech Inc. Stamford, Connecticut **1991,** PD-4.

[12] Spectra-Tech Product Data Sheet, *ATR Microscopy*, Spectra-Tech Inc. Stamford, Connecticut **1991,** PD-8.

[13] C. J. Curry, M. J. Whitehouse, J. M. Chalmers, *Appl. Spectrosc.* **1985,** *39*, 174.

[14] J. M. Chalmers, M. W. Mackenzie in *Advances in Applied Fourier Transform Infrared Spectroscopy* (Ed.: M. W. Mackenzie), John Wiley, Chichester, UK **1988,** Chapter 4.

[15] M. Sekkal, J. P. Huvenne, B. Sombret, P. Legrand, M. C. Verdus, J. C. Mollet, A. Moradi, *Mikrochim. Acta* **1993,** *112*, 1.

[16] E. G. Bartick, M. W. Tungol, J. A. Reffner, *Anal. Chim. Acta* **1994,** *288*, 35.

Part II

X-Ray Microscopy

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

1 Soft X-Ray Imaging

1.1 Introduction

Soft X rays in the wavelength range of about 0.3–5 nm are very well suited for X-ray microscopy studies. The reasons are that the interaction of soft X-rays with matter is very favorable for such studies and that high-resolution X-ray optical elements can be built for this wavelength range.

The interaction of soft X rays with matter is dominated by elastic scattering and photoelectric absorption, or, expressed in a different way, by a phase shift of the incoming wave and an attenuation of its amplitude.

We consider a material of thickness t and refractive index

$$\tilde{n} = n - i\beta = 1 - \delta - i\beta \tag{1}$$

illuminated by a plane wave with the amplitude $\mathscr{A}_0$. If $\mathscr{A}_1$ is the amplitude of the transmitted wave, the amplitude transmission is given by

$$\begin{aligned}\mathscr{T} &= \frac{\mathscr{A}_1}{\mathscr{A}_0} \\ &= \exp\left(-\frac{2\pi}{\lambda}\beta t\right)\exp\left(i\frac{2\pi}{\lambda}\delta t\right) \\ &\quad \times \exp\left(-i\frac{2\pi}{\lambda}t\right)\end{aligned} \tag{2}$$

The first factor describes the attenuation of the amplitude of the incoming wave, the second factor describes the phase shift of the incoming wave with respect to a wave in vacuum. With $\mathscr{A}_0\mathscr{A}_0^* = I_0$ and $\mathscr{A}_1\mathscr{A}_1^* = I_1$ one gets

$$\frac{I_1}{I_0} = \exp\left(-\frac{4\pi}{\lambda}\beta t\right) \tag{3}$$

where $4\pi\beta/\lambda$ is called the linear absorption coefficient μ_1. If one divides μ_1 by N, that is, by the number of atoms per cubic centimeter, one gets the atomic cross-section for photoelectric absorption $\sigma_a = \mu_1/N$. Tabulations of X-ray absorption data often give the mass absorption coefficient $\mu = \mu_1/\rho$.

To describe the phase shift, the amplitudes must be considered instead of the intensities. In analogy to the linear absorption coefficient one can introduce a phase shift coefficient, namely $\eta = 2\pi\delta/\lambda$.

Phase shift and absorption, described by the complex index of refraction, can be attributed to the scattering of the incoming wave by the atoms of the material, described by the atomic scattering factor $f = f_1 + if_2$ [1]. The amplitude scattered by an atom is given by this atomic scattering factor f multiplied by the amplitude which would be scattered if the atom is replaced by a free electron. Superposition of the

amplitudes scattered at the single atoms of the material gives a resultant wave. So the interaction is described by f_1 and f_2. The relation between the complex index of refraction and the atomic scattering factor is given by [1]

$$\delta = \frac{r_0 \lambda^2}{2\pi} N f_1 \tag{4}$$

and

$$\beta = \frac{r_0 \lambda^2}{2\pi} N f_2 \tag{5}$$

where $r_0 = 2.82 \times 10^{-15}$ m is the classical electron radius, $N = N_A \rho / M$ is the number of atoms, N_A is Loschmidt's constant, M is the molar mass, and ρ is the density of the material. The atomic scattering factor has been tabulated for 94 elements for the wavelength range $0.04\,\text{nm} \leqslant \lambda \leqslant 24.8\,\text{nm}$ [2]. f_1 and f_2 vary very strongly with photon energy and chemical composition, and so do the linear absorption coefficient μ_1 and the phase shift coefficient η.

For soft X rays, microscopic specimens are amplitude objects as well as phase objects, and X-ray microscopy studies can be performed in amplitude contrast and in phase contrast. For biological studies performed in amplitude contrast the wavelength range between 2.3 and 4.4 nm is best suited, that is, between the K edges of oxygen and carbon. In Fig. 1 the linear absorption coefficient μ_1 of water and a protein with the composition $C_{94}H_{139}N_{24}O_{31}S$ and the density $\rho = 1.35\,\text{g cm}^{-3}$ is shown. The figure demonstrates that in this wavelength range, water is much more transparent than the organic material. Hence it follows that a natural contrast is observed for cells and cell organelles in a natural, wet state. Staining to increase the contrast is not necessary for X-ray microscopy studies. But it can of course be advantageous to use specific staining. The high transparency of water means that layers with a thickness of up to about 10 μm can be irradiated. The wavelength range mentioned above is therefore called the water window.

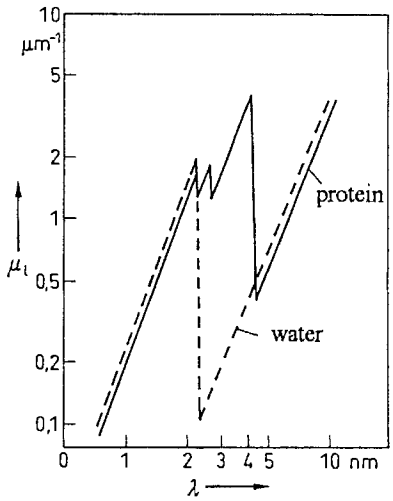


Figure 1. Linear absorption coefficient μ_1 for a protein and water as a function of wavelength.

X-ray micrographs obtained in amplitude contrast have the advantage of the ease of image interpretation. For chemically unfixed hydrated biological specimens, however, it can be advantageous to use a phase contrast because a considerably higher image contrast can be reached as shown in Sec. 2 of this Chapter.

From the f_1 and f_2 values, tabulated by Henke et al. [2], it follows that δ and β of the complex refraction index are much smaller than 1 in the whole range of soft X radiation. This property of X rays causes a near absence of reflections at interfaces and results in very clear images, even of thick specimens, in many X-ray microscopy applications. On the other

hand, the small δ and β values make it more difficult to build high-resolution X-ray optical elements.

For carbon and the wavelength $\lambda = 3.5\,\text{nm}$, for example, one gets $\delta = 3 \times 10^{-3}$ and $\beta = 2 \times 10^{-3}$. For the linear absorption coefficient it follows that $\mu_1 = 7\,\mu\text{m}^{-1}$. If one wishes to construct a convergent refractive X-ray lens for an X-ray microscope, it would have to be very thin because of the strong absorption. Such a plane-concave resp. biconcave lens—because of $n < 1$—would have a rather long focal length and a small numerical aperture. So, refractive lenses for soft X rays cannot be used in practice in X-ray microscopes.

The other conventional method for imaging is to use mirrors. The reflectivity ρ can be calculated with the Fresnel formula, and results in $\rho \leqslant 10^{-5}$ in the soft X-ray range for metal layers used in normal incidence. The reflectivity of mirrors is enhanced considerably for X rays impinging with grazing incidence. Wolter [3] has proposed two-mirror systems for grazing incidence as high-resolution objectives. But because of the necessary high quality of the mirrors and a restricted object field, such optical elements have not been very succesful up to now for high-resolution X-ray microscopy.

Another possibility to enhance the reflectivity is to use multilayer mirrors, where the refractive index varies periodically with depth. To build such mirrors, alternate layers are evaporated or sputtered onto a blank with a thickness of $\lambda/4$ and relatively high refractive index n_1 and a relatively low refractive index n_2, respectively [4]. The weak radiation reflected at the interfaces of the multilayer is superimposed coherently and in phase, and can give a considerably increased reflectivity. At present, sufficiently high reflectivities in normal incidence have only been reached for wavelengths $\lambda \geqslant 5\,\text{nm}$.

Up to now, X-ray microscopy experiments have been performed with mirrors in grazing incidence, with multilayer mirrors in normal incidence and with zone plates, that is, diffractive X-ray lenses. The latter give at present by far the best resolution at wavelengths suitable for X-ray microscopy. One reason for this is that aberrations scale with the size of the X-ray optical elements and that zone plates can be built as micro zone plates with diameters in the range of 50 μm. Therefore, in Secs. 2 and 3 of this Chapter only X-ray microscopes will be discussed which contain zone plates as high-resolution X-ray lenses.

Zone plates are diffractive focusing elements and consist of alternating transparent and absorbing phase-shifting circular zones. Their focusing properties are well known [5]. Imaging with zone plates with zone numbers $n \geqslant 100$ obeys the same laws as imaging with thin refractive lenses. With r_1 the radius of the innermost zone, r_n the radius of the nth zone, and n the zone number, the focal length f of a zone plate used in the mth diffractive order is approximately given by $f = r_1^2/m\lambda$. In the following only the first diffractive order $m = 1$ is considered. For optimum performance a zone plate should be used with quasimonochromatic radiation with $\lambda/\Delta\lambda \approx n$. As zone plates work like thin lenses, they have a diffraction limit of $\Delta = 0.61\,\lambda/\text{NA}$. The numerical aperture NA is given by $\lambda/2\,\text{d}r_n$, where $\text{d}r_n$ is the width of the outermost zone. The resolution is therefore given by

$\Delta = 1.22\,dr_n$, and is approximately equal to the width of the outermost zone.

An important parameter characterizing a zone plate is the efficiency, that is, the ratio of the intensity in the focus of the first order divided by the intensity of the incident radiation. For alternating opaque and transparent zones, the first-order diffraction efficiency is about 10% for optically thin zone plates [6]. If the opaque zones are replaced by a material that introduces a phase shift of π, the efficiency can be increased theoretically to about 40%. But as no material exists in the soft X-ray region which shows a phase shift without absorption, an efficiency of only about 20% can be reached. For high-resolution zone plates, the thickness of the zones becomes much larger than the zone width. Such high aspect ratio nanostructures start to act as waveguides for soft X rays. Zone plates with these nanostructures, for example germanium or nickel zones with $dr_n < 20\,\text{nm}$ and aspect ratios >10, can no longer be treated as thin gratings for soft X rays. The diffraction efficiency of such zone plates, calculated with a dynamical theory, can be considerably larger than the efficiency of thin zone plates, namely more than 40% [7]. For X-ray microscopes, two types of zone plate are needed. First, the object has to be illuminated by a condenser which has to collect as much radiation as possible from the X-ray source. For X-ray microscopy with synchrotron radiation from electron storage rings, condensers with diameters of up to 9 mm and zone number $n = 4.2 \times 10^4$ are used. Secondly, to image the object, high-resolution zone plates are necessary, that is, zone plates with outermost zone widths dr_n as small as possible. These zone plates have small diameters—as already mentioned—to limit the geometric-optical aberrations, and have only a few hundred zones in order not to restrict the usable X-ray bandwidth too much.

Two methods of zone plate fabrication have been used up to now. Zone plate patterns are generated by the superposition of two coherent, aspheric ultraviolet wavefronts [8]. This method is especially suited for large condenser zone plates. The interference pattern is recorded in a photoresist layer and transferred into absorbing and phase shifting zones by reaction ion etching using tri-level pattern transfer techniques [9].

At present, the best method for the construction of high-resolution zone plates is electron beam lithography. With this technique the zone plate pattern is written into an electron beam resist, polymethylmethacrylate (PMMA) with a small electron beam [10–12]. The resist pattern is then transferred into, for example, germanium or nickel zones with high aspect ratios by use of tri-level pattern transfer techniques. The smallest zone width in germanium that has been obtained up to now is 19 nm [13].

1.2 Transmission X-Ray Microscopy

X-ray microscopes are in operation and under development all over the world. Because synchrotron radiation sources have at present the highest spectral brilliance in the soft X-ray range, most of the instruments are installed at electron storage rings. In addition, X-ray microscopes are under development which use

pulsed plasma sources as laser-produced plasmas or gas plasma discharge sources. The state of the art of transmission X-ray microscopy (TXM) and scanning transmission X-ray microscopy (STXM) as well as X-ray microscopy studies performed with these techniques are described in several conference proceedings [14–20] and review articles [21, 22].

In this contribution, two X-ray microscopes installed at electron storage rings are described, namely the Göttingen X-ray microscope at the synchrotron radiation source BESSY in Berlin, FRG, and the Stony Brook scanning X-ray microscope at the National Synchrotron Light Source at Brookhaven National Laboratory in the USA.

Figure 2 shows the X-ray optical schematic of the X-ray microscope at BESSY [23]. The continuous synchrotron radiation is incident on a condenser zone plate which focuses X radiation of the wanted wavelength into a pinhole placed in the object plane. Due to the wavelength dependence of the focal length of a zone plate, the condenser acts in combination with the pinhole as a linear monochromator with a spectral resolution of $\lambda/d\lambda = D/2d$. D is the diameter of the condenser zone plate and d is the diameter of the pinhole. So quasimonochromatic radiation is used for imaging which is necessary to get diffraction limited resolution. The central stop of the condenser prevents the object being illuminated by broad-band zero-order radiation from the condenser. In addition, due to the stop, the image field is free of zero-order radiation from the micro zone plate. This high-resolution X-ray lens downstream of the object generates a magnified X-ray image of the object. This image is recorded with a charge-coupled device (CCD) detector. In addition, the image can be made visible with a channel plate detector for low-resolution real-time observation to adjust the X-ray optical set-up.

For phase contrast X-ray microscopy experiments the X-ray optical arrangement as shown in Fig. 2 is supplemented by an annular aperture in front of the condenser and an annular phase plate in the back focal plane of the micro zone plate [24], which corresponds to a phase contrast microscope for visible light.

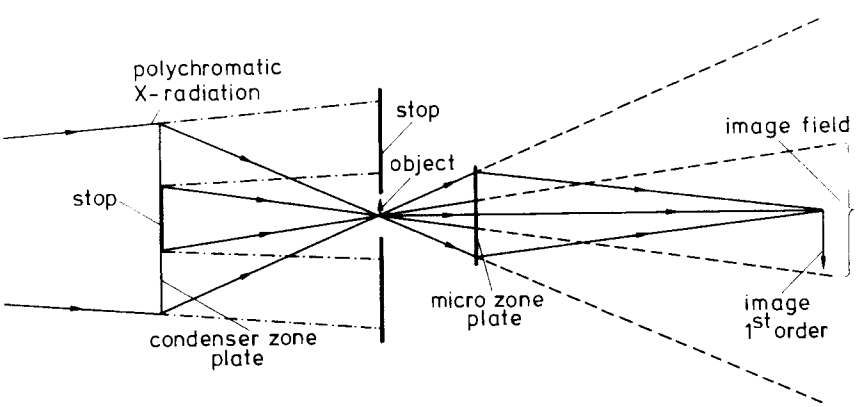


Figure 2. X-ray optical schematic of an X-ray microscope with zone plate optics.

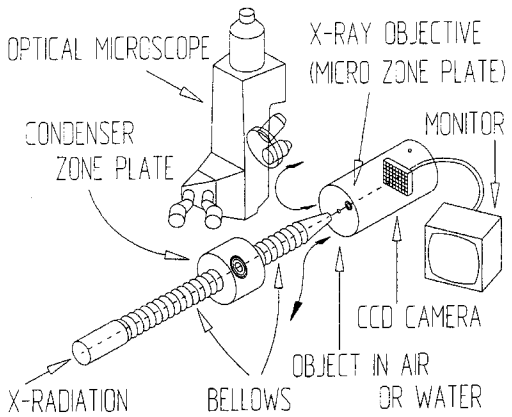


Figure 3. Schematic of the X-ray microscope at the electron storage ring BESSY, Berlin.

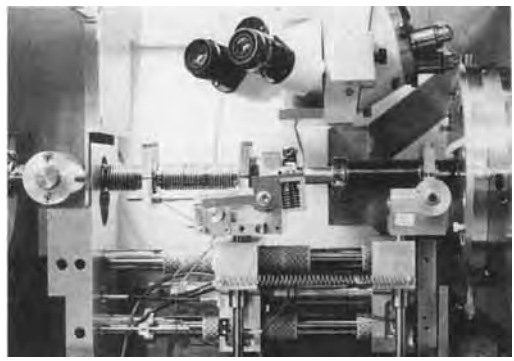


Figure 4. Central part of the X-ray microscope at the electron storage ring BESSY, Berlin.

As the range of soft X rays in air is only a few millimeters, the major part of the X-ray optical path is in vacuum. Figure 3 shows a schematic diagram of the X-ray microscope [25]. The condenser zone plate is mounted in a bellows which ends in a thin foil transparent to soft X rays. This foil separates the vacuum of the bellows from air. The micro zone plate and the CCD camera are arranged in a second vacuum chamber. The object is mounted in an environmental object chamber where it can be prepared between two thin foils in its natural state, including when wet. The object chamber is in air and easily accessible. The bellows with the condenser can be bent downward to swivel an optical microscope into position, which allows adjustment of the object and prefocusing of the object before X-ray imaging.

The CCD camera contains a thinned, back-illuminated CCD chip with 1024×1024 pixels [26]. It combines high spatial resolution, a large dynamic range and good linearity with high values of detective quantum efficiency, for example 70% at $\lambda = 2.5\,\text{nm}$. The image can be read out electronically, which allows the possibility of direct quantitative measurements, image processing, etc. Figure 4 shows a photograph of the central part of the X-ray microscope.

For an image of a hydrated specimen with a resolution in the range of 25–50 nm taken with X rays of $\lambda = 2.5\,\text{nm}$, an exposure time of 1–10 s is needed. With these parameters, X-ray microscopy studies have been performed in different fields, such as biology, biophysics, colloid chemistry, and soil sciences. Some examples are described below.

Figure 5 shows a pair of images taken in amplitude and phase contrast. The object is a Kupffer cell of rat liver from a primary culture in a hydrated state, fixed in glutaraldehyde. The images were taken to identify macrophage-specific antigens on the surface of the Kupffer cells using specific antibodies and silver staining [27]. Figure 5a shows such a cell imaged in amplitude contrast. The stained dark areas can be localized very well. Figure 5b shows the same cell imaged in phase contrast. Many more fine structures can be detected in this image, in the nucleus as well as in the cytoplasm, whereas the stained dark areas

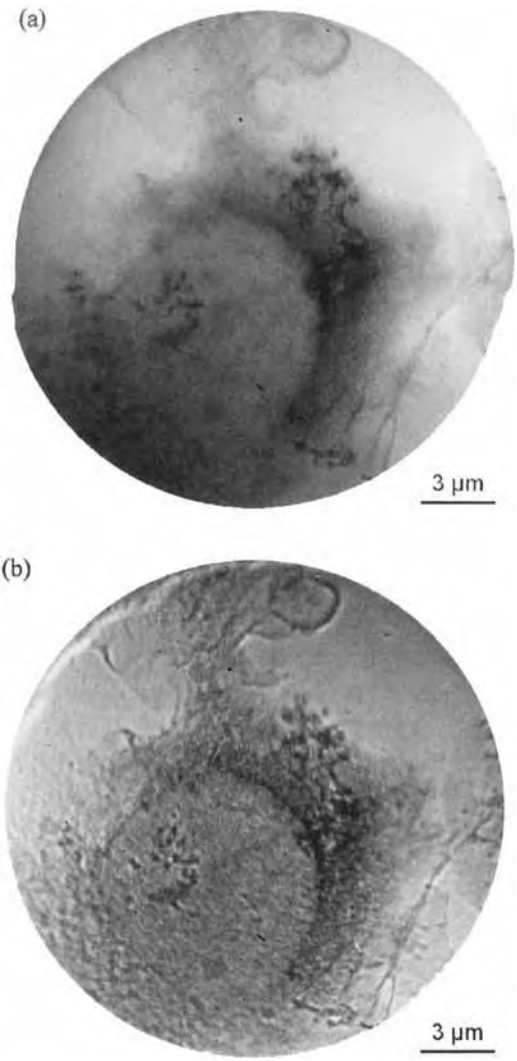


Figure 5. (a) Kupffer cell of the liver of the rat from a primary culture in a hydrated state, fixed in glutaraldehyde, $\lambda = 2.4$ nm, using amplitude contrast. (b) Same object as in part (a), but with phase contrast.

(compare Fig. 5a) are less emphasized. This example shows that it may be advantageous to take images using both amplitude and phase contrast.

The sample used for the micrograph in Fig. 6 was a 0.1% suspension of Na^+ montmorillonite clay in a culture medium

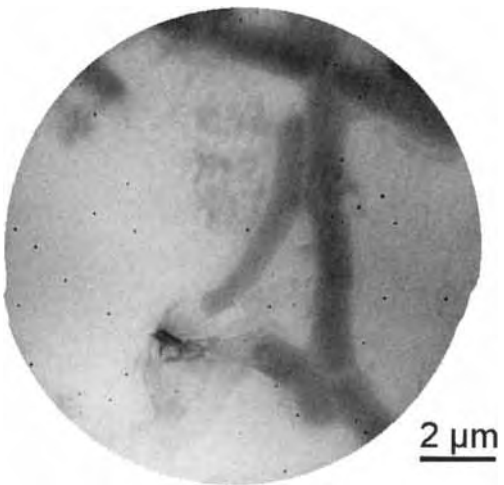


Figure 6. Suspension of montmorillonite clay in a culture medium with the bacterium *Bacillus megaterium*, $\lambda = 2.4$ nm. (The image was taken by J. Thieme, J. Niemeyer, G. Machulla, and P. Guttmann.)

with the bacterium *Bacillus megaterium*. The image shows several bacteria surrounded by their slime envelopes (mucinous gel), which interact with the clay particles. A small clay aggregate is totally embedded by an envelope. As the envelope consists largely of water, such images with the required resolution can only be taken with X-ray microscopes.

Three-layer clay minerals are able to swell when in contact with water, so enlarging their surfaces considerably. The swelling and thus the delaminating of the stacks of clay particles can be seen in the X-ray micrograph of Fig. 7, which shows a stack of the three-layer clay mineral nontronite in water. Such structures can only be examined with X-ray microscopes, as the brush-like ends of the stacks seen in Fig. 7 would collapse when imaged in a dry environment.

For biological objects the absorption of soft X rays can cause structural changes, especially in rather sensitive vertebrate

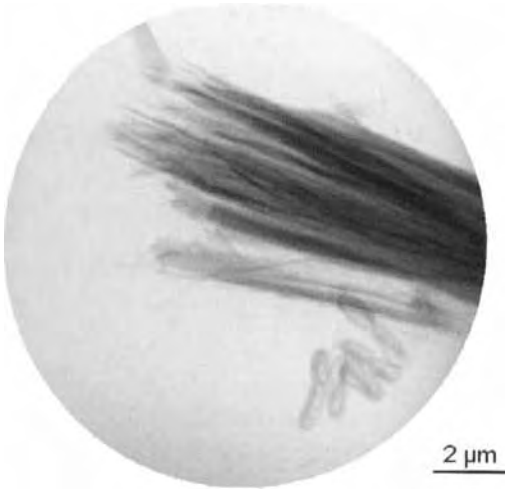


Figure 7. The three-layer clay mineral nontronite in water, $\lambda = 2.4$ nm. (The image was taken by J. Thieme, J. Niemeyer, and P. Guttmann.)

cells. These changes are due to radiation-induced kinetics and the transport of matter by diffusion, and become important in the dosage range 10^4–10^6 Gy for unfixed cells and cell organelles. For cells fixed with glutaraldehyde the tolerable dosage is about 10^5–10^6 Gy, which is the dosage necessary to take an image like Fig. 5. Theoretical investigations have shown that chemically unfixed frozen–hydrated specimens should be structurally stable up to dosages of about 10^9 Gy. A cryogenic object chamber has been developed and implemented at the X-ray microscope at BESSY [28]. Chemically unfixed algae frozen at about 120 K were irradiated over a long time corresponding to a dosage of up to 10^{10} Gy. No structural changes were observed within the resolution limit of 50 nm. The cryogenic method will allow imaging of chemically unfixed specimens with a resolution of less than 10 nm. Furthermore, the high stability of frozen–hydrated objects will allow to take multiple images for three-dimensional image formation, for example in X-ray microscopic tomography.

1.3 Scanning Transmission X-Ray Microscopy

The primary components of a scanning X-ray microscope are a microspot-forming optic, a scanning stage, a detector, and a computer control system. Figure 8 shows the schematic of a scanning X-ray microscope installed by Kirz and co-workers [22, 29] at the National Synchrotron Light Source at Brookhaven National Laboratory, USA. The microspot is formed by illuminating a zone plate with spatially coherent quasimonochromatic X rays. To isolate the beam diffracted into the first-order focus, the zone plate has a central stop. In addition, an order-selecting aperture is inserted between the zone plate and the focal plane. The specimen is mounted in an object chamber which can be scanned in two orthogonal directions with stepper motors and piezos, respectively. The radiation transmitted through the object is detected point by point by a proportional counter, and the X-ray image is displayed on a monitor. To get diffraction-limited resolution given by the micro zone plate, one has to use spatially coherent radiation. That means that X-ray sources with high spectral brilliance, such as the undulators or wigglers of electron storage rings, should be used. The operator of the microscope is free to select the number of pixels in the image (up to 500×500), the pixel size, and the dwell time for each pixel. A typical large-scale

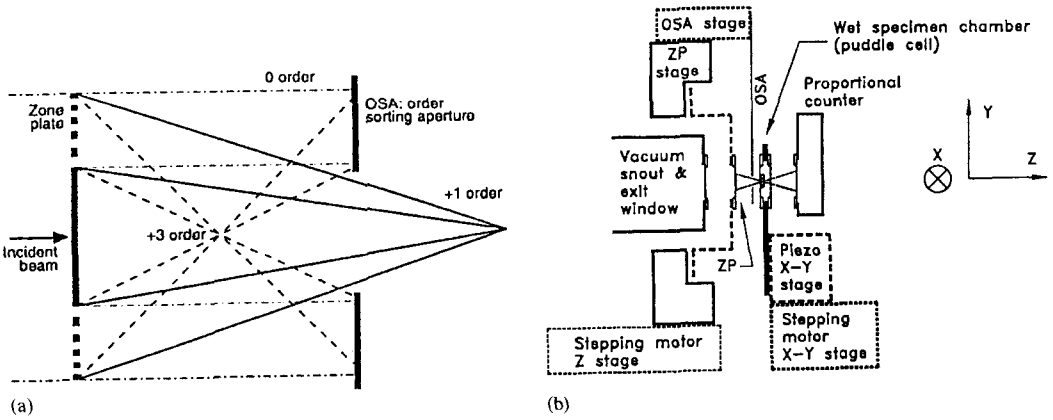


Figure 8. Schematic of the scanning X-ray microscope at the National Synchrotron Light Source at Brookhaven National Laboratory. (a) Arrangement of the zone plate and order-selecting aperture (OSA). (b) The specimen is placed on an X/Y scanning stage with coarse and fine motions driven by stepping motors and piezos, respectively. The image is focused by moving the zone plate and the OSA together as a unit in Z. The distance from the zone plate (ZP) to the OSA is $\approx$0.7 mm at $\lambda \approx 4$ nm, and the distance from the OSA to the specimen is $\approx$0.3 mm. (Courtesy of C. Jacobsen, State University of New York at Stony Brook, USA.)

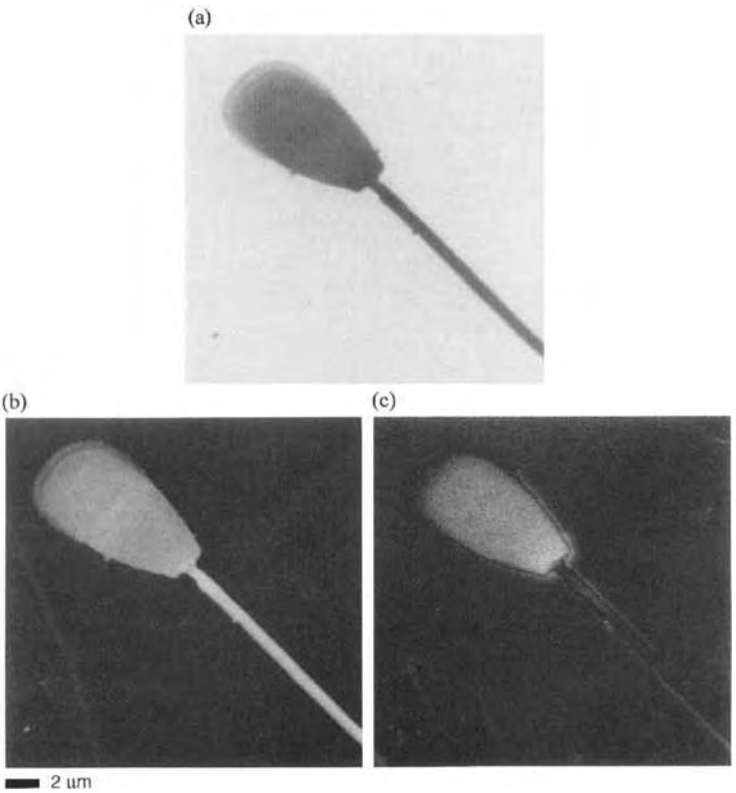


Figure 9. (a) STXM image ($\lambda = 4.268$ nm), and (b) protein and (c) DNA maps, of air-dried bull sperm (prepared by R. Balhorn, Lawrence Livermore National Laboratory). The images were taken in the NEXAFS resonance spectral region near the carbon K edge. (Courtesy of X. Zhang, State University of New York at Stony Brook, USA.)

orientation scan involves 400 × 400 pixels with 1 μm steps and a 1 ms dwell time using the stepper motors. For a high-resolution image the piezos are used for scanning with a pixel size of about 50 nm. A 300 × 300 pixel scan with a 5 ms dwell or about 5000 incident photons per pixel will take about 8 min and impart a dose of about 4×10^5 Gy to the sample [22].

The scanning X-ray microscope is being used for imaging of hydrated specimens, for elemental and chemical mapping, for luminescence studies [22], and as a scanning photoemission microscope for surface analysis [30]. Here, one example is shown where the scanning microscope has been used for chemical mapping. Different molecules, for example DNA or proteins, show strong differences in the near-edge X-ray absorption fine structure (NEXAFS) spectra near the carbon K edge, that is, the linear absorption coefficient as a function of wavelength shows absorption peaks with differences in peak intensities as well as in peak positions, caused by resonances from the double bonds C=C, C=N, and C=O in the molecules. So the DNA and protein compounds can be distinguished by taking images at different absorption maxima. From these absorption spectra, protein and DNA maps can be derived. Figure 9 shows the DNA and protein maps of an air-dried bull sperm. The image at the top, taken at $\lambda = 4.268$ nm, shows the contrast due to all-carbon absorption [31].

1.4 References

[1] R. W. James, *The Optical Principles of the Diffraction of X-Rays*, Ox Bow Press, Woodbridge, CT **1982**.

[2] B. L. Henke, E. M. Gullikson, J. C. Davis, *At. Data Nucl. Data Tables* **1993,** *54*, 181.

[3] H. Wolter, *Ann. Physik*, **1952,** *6*(10), 10.

[4] E. Spiller, *Appl. Phys. Lett.* **1972,** *20*, 365.

[5] J. L. Soret, *Ann. Phys. Chem.* **1875,** *156*, 99.

[6] J. Kirz, *J. Opt. Soc. Am.* **1974,** *64*(3), 301.

[7] J. Maser in *X-Ray Microscopy IV* (Eds.: V. V. Aristov, A. I. Erko), Bogorodskii Pechatnik, Chernogolovka, Russia **1994,** p. 523.

[8] G. Schmahl, D. Rudolph, P. Guttmann, D. Christ in *X-Ray Microscopy* (Eds.: G. Schmahl, D. Rudolph), Springer, Berlin **1984,** p. 63.

[9] T. Matthies, C. David, J. Thieme, *J. Vac. Sci. Technol. B* **1993,** *11*(5), 1873.

[10] V. Bögli, *Proc. SPIE* **1986,** *733*, 449.

[11] E. H. Anderson, D. Kern in *X-Ray Microscopy III* (Eds.: A. G. Michette et al.), Springer, Berlin **1992,** p. 75.

[12] C. David, J. Thieme, P. Guttmann, G. Schneider, D. Rudolph, G. Schmakl, *Optik* **1992,** *91*(3), 95.

[13] G. Schneider, T. Schliebe, H. Aschoff, *J. Vac. Sci. Technol. B* **1995,** 2809.

[14] G. Schmahl, D. Rudolph (Eds.), *X-Ray Microscopy*, Springer, Berlin **1984.**

[15] P. C. Cheng, G. J. Jan (Eds.), *X-Ray Microscopy—Instrumentation and Biological Applications*, Springer, Berlin **1987.**

[16] D. Sayre, M. Howells, J. Kirz, H. Rarback (Eds.), *X-Ray Microscopy II*, Springer, Berlin **1988.**

[17] K. Shinohara, K. Yada, H. Kihara, T. Saito (Eds.), *X-Ray Microscopy in Biology and Medicine*, Japan Scientific Societies, Tokyo/Springer, Berlin **1990.**

[18] A. G. Michette, G. R. Morrison, C. J. Buckley (Eds.), *X-Ray Microscopy III*, Springer, Berlin **1992.**

[19] C. J. Jacobsen, J. E. Trebes (Eds.), *Proc. SPIE* **1992,** *1741.*

[20] V. V. Aristov, A. I. Erko (Eds.), *X-Ray Microscopy IV*, Bogorodskii Pechatnik, Chernogolovka, Russia **1994.**

[21] G. Schmahl, P. C. Cheng in *Handbook on Synchrotron Radiation*, Vol. 4 (Eds.: S. Ebashi et al.), Chap. 15, North-Holland, Amsterdam **1991,** p. 481.

[22] J. Kirz, C. Jacobsen, M. Howells, *Q. Rev. Biophys.* **1995,** *28*, 33.

[23] D. Rudolph, B. Nicmann, G. Schmahl, D. Christ in *X-Ray Microscopy* (Eds.: G. Schmahl, D. Rudolph), Springer, Berlin **1984,** p. 192.

[24] G. Schmahl, D. Rudolph, P. Guttmann, G. Schneider, J. Thieme, B. Niemann, *Rev. Sci. Instrum.* **1995,** *66*(2), 1282.

[25] B. Niemann, G. Schneider, P. Guttmann, D. Rudolph, G. Schmahl in *X-Ray Microscopy IV* (Eds.: V. V. Aristov, A. I. Erko), Bogorodskii Pechatnik, Russia **1994,** p. 66.

[26] T. Wilhein, D. Rothweiler, A. Tusche, F. Scholze, W. Meyer-Ilse in *X-Ray Microscopy IV* (Eds.: V. V. Aristov, A. I. Erko), Bogorodskii Pechatnik, Russia **1994,** p. 470.

[27] G. Schmahl, P. Guttmann, G. Schneider, B. Niemann, C. David, T. Wilhein in *X-RayMicroscopy IV* (Eds.: V. V. Aristov, A. I. Erko), Bogorodskii Pechatnik, Russia **1994,** 196.

[28] G. Schneider, B. Nicmann, P. Guttmann, D. Rudolph, G. Schmahl, *Synchrotron Radiation News* **1995,** *8*(3), 19.

[29] C. Jacobsen, E. Anderson, H. Chapman, J. Kirz, S. Lindaas, M. Rivers, S. Wang, S. Williams, S. Wirick, X. Zhang in *X-Ray Microscopy IV* (Eds.: V. V. Aristov, A. I. Erko), Bogorodskii Pechatnik, Russia **1994,** p. 304.

[30] H. Ade, J. Kirz, S. L. Hulbert, E. Johnson, E. Anderson, D. Kern, *App. Phys. Lett.* **1990,** *56*, 1841.

[31] X. Zhang, R. Balhorn, C. Jacobson, J. Kirz, S. Williams in *Proc. 52nd Annual Meeting Microscopy Society of America* (Eds.: G. W. Bailey, A. J. Garrat-Reed), San Francisco, San Francisco **1994,** p. 50.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

2 X-Ray Microradiography

2.1 Physical Principles

2.1.1 Interaction Processes Involved in Microradiography

X-ray microradiography (XM), which includes X-ray contact microradiography (XCM) and X-ray projection microradiography (XPM), is concerned with X-ray imaging systems that do not require any kind of optical device such as pinhole, mirror, or lens [1, 2]. Contrast in images results from the attenuation of X-rays in the low or medium energy range (0.1–50 keV) by passing through the specimen. This attenuation is governed by Beer's law

$$I = I_0 e^{-\mu(E)t} \tag{1}$$

where I_0 and I are the incident and transmitted X-ray intensities, respectively, t is the thickness of the specimen, and $\mu(E)$ is the linear absorption coefficient of the absorber for X-ray energy E.

The attenuation of X-rays through matter is characterized by a total attenuation cross section σ. Because the attenuation process is a result of atomic interactions, the linear absorption coefficient for a compound follows a simple law of weighted averages. Then, if the absorber is composed of elements A, B, ..., P with atomic densities N_A, N_B, ..., N_P (at. cm^{-3}), the linear absorption coefficient is related to the total attenuation cross section of the elements σ_A, σ_B, ..., σ_P (cm^2) by the relation

$$\mu(E) = N_A\sigma_A(E) + N_B\sigma_B(E) + \ldots + N_P\sigma_P(E) \tag{2a}$$

or, in terms of the atomic concentrations C_A, C_B, ... (at.%), by the relation

$$\mu(E)t = \frac{N^0 t}{100}\sum_j C_j\sigma_j \tag{2b}$$

where N^0 is the total atomic density (about $5 \times 10^{22}\ cm^{-3}$ for most solids).

The total attenuation cross section for a given element, say A, can be expressed as the sum of three coefficients, each of them representing the contribution of the different processes involved in attenuation. Thus,

$$\sigma_A = \sigma_A^p + \sigma_A^i + \sigma_A^c \tag{3}$$

σ_A^p is the photoelectronic absorption cross section associated with the photoelectric effect, σ_A^i is the incoherent scattering cross section associated with the Compton effect, σ_A^c is the coherent scattering associated with Rayleigh scattering. A more in-depth treatment is given in [3].

In the photon energy range for X-ray microradiography (0.1–50 keV), as shown

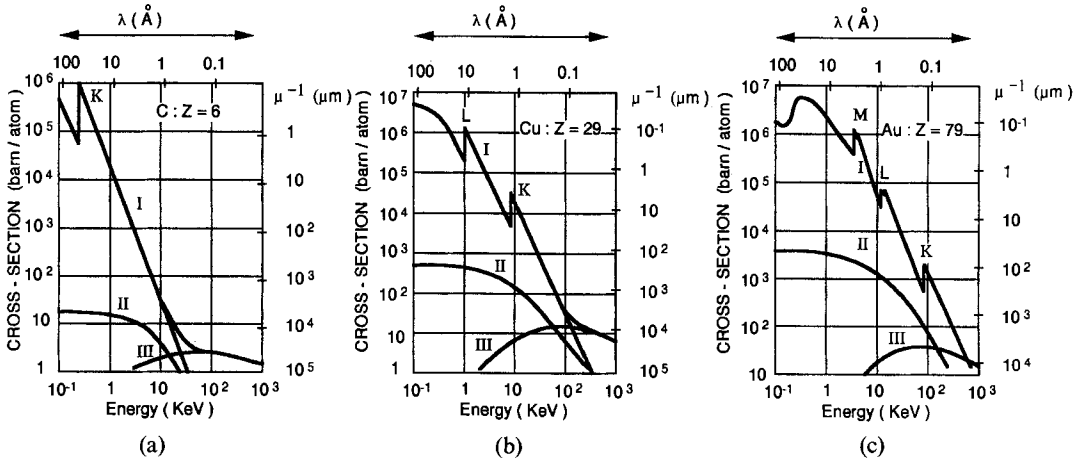


Figure 1. X-ray attenuation cross sections for carbon (a), copper (b), and gold (c) after Veigele [4]. For each of the elements the contributions of σ^p (I), σ^i (II), and σ^c (III) are shown.

in Fig. 1, the most probable interaction of X-ray photons with matter is the photoelectric effect (i.e., when an inner electron of the atom is removed from its shell by the incoming photon). The cross section, σ, can be obtained from tables [4–8], calculated [9, 10] or measured using standards, and the linear absorption coefficients μ (cm^{-1}) estimated as well as the mass absorption coefficients μ/ρ ($cm^2\,g^{-1}$).

2.1.2 Specimen Thickness and Sensitivity of X-Ray Microradiography

The ability of X-ray microradiography to image a specimen with good contrast depends on differences in absorption between two neighboring areas of this specimen. For a given photon energy, cross sections strongly increase with the atomic number. This can be expressed as

$$\sigma \simeq Z^n \quad (1 < n < 4) \tag{4a}$$

This explains the high Z-contrast obtained on X-ray absorption images. For example, biological samples need not be stained with heavy materials to enhance contrast.

For a given element, away from any absorption edges, the cross section decreases with increase in photon energy. Thus,

$$\sigma \simeq (h\nu)^{-3} \tag{4b}$$

A typical thickness, t, for a specimen to be investigated by X-ray absorption technique is of the order $1/\mu$; specimen thicknesses increase rapidly with the photon energy applied.

High penetration capability of X-rays is the major advantage of X-ray microscopy over most of the other microscopies; hence, a vacuum is not required and the specimen can be set in air, water or any kind of selected environment.

From Eq. (4b) it is clear that the choice of an incident X-ray energy for imaging is critical to the contrast of the final image. At absorption edges, a discontinuity occurs in the absorption curve $\sigma(E)$

(Fig. 1) which provides a unique method for element discrimination. Thus, the weight per unit area of a given element can be deduced from two X-ray absorption images taken at two energies above and below the edge [11–13].

The ability of X-ray absorption microradiography to detect faint features in the object is dependent on the number of incident photons per picture element (pixel), n_0. By applying Rose's criterion, which states that for a given relative variation $\Delta\mu t/\mu t$ to be detectable in the image it must stand out with a signal that is at least three times greater than the standard deviation of the surrounding background [14–16], we have

$$\frac{\Delta\mu t}{\mu t} = \frac{3}{\sqrt{n_0}} \times \frac{e^x}{2x} \tag{5}$$

where $x = \mu t/2$.

An $e^x/2x$ against x plot (Fig. 2) shows that the best sensitivity with XM is obtained when the thickness of the specimen is related to the linear absorption coefficient by the relation $\mu t_0 = 2$. In this case, we calculate that the required number of photons to detect a 1% relative change of μt is 10^5.

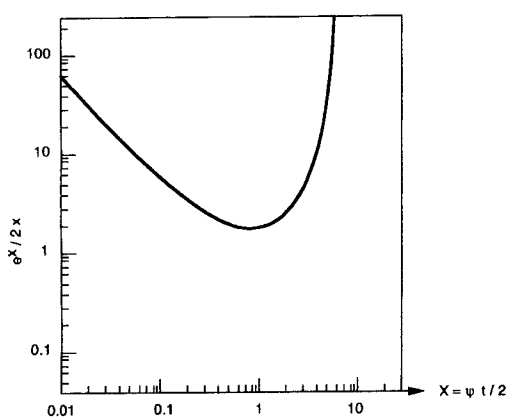


Figure 2. Sensitivity in XCM according to Eq. (5). The best sensitivity is obtained when $\mu t_0 = 2$.

The minimum atomic concentration C_A (at. %) for a given element A, of cross section σ_A, embedded in a matrix M of cross section σ_M, to be detected is given by

$$C_{A\,\min} \simeq \left(\frac{\Delta\mu t}{\mu t}\right)\left(\frac{\sigma_M}{|\sigma_A - \sigma_M|}\right) \times 100 \tag{6}$$

This result shows that the detection of cracks or empty cavities in a specimen ($\sigma_A = 0$) is governed by the corresponding change of the local thickness of the specimen. A more favorable situation occurs when the cross section for the feature σ_A is far larger than the cross section of the matrix. For 10 keV photons, the detection limit of gold or copper in a carbon matrix a few millimeters thick may reach 10^{-3} at. %.

2.2 Contact Microscopy

2.2.1 Principle

X-ray contact microscopy (XCM), commonly referred to as X-ray contact microradiography, was the very first technique used to obtain X-ray images on a microscopic scale; the first attempts using this technique took place in the early years when Röntgen discovered X-rays in 1895 [1, 2]. It is, in principle, the simplest method to implement and, paradoxically, it is also the only method to approach the theoretical limit of resolution according to Fresnel diffraction.

A microradiography image is obtained by placing a specimen in close contact to an

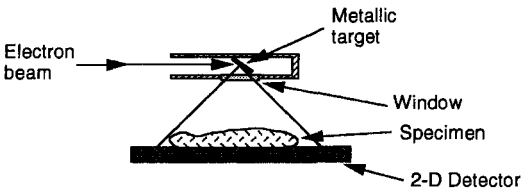


Figure 3. Basic apparatus for X-ray contact microscopy.

appropriate X-ray sensitive medium and exposing it to X-rays. The experimental layout is depicted in Fig. 3. A latent image results from shadow casting on the recording plane (e.g., photographic plates). After processing, the image exhibits a contrast locally dependent on the absorption of the X-ray beam during its passage from source to detector through the specimen. Because the magnification is unity, a further enlargement of the image, by optical microscopy or another microscopy technique, is needed.

In the 1960s, after a period of major development, research activities in X-ray microscopy tapered off, to be superseded by electron microscopy. A renewal of interest in XCM began in the 1980s with the development of high resolution X-ray photosensitive materials, known as photoresists, and with the advent of new kinds of intense X-ray sources, such as synchrotron radiation or plasma sources. Most of the recent applications of XCM concern the soft X-ray range (SXCM).

2.2.2 Resolution

Three main factors limit resolution in XCM: (i) the thickness of the specimen, t; (ii) the finite size of the source, s; and (iii) the ability of the recording medium to resolve fine features.

The size of the source (s) induces a penumbral blurring whose width is given by

$$\delta_p \cong \frac{st}{d} \tag{7}$$

where d is the distance from the source to the specimen front plane (Fig. 4a). Penumbral blurring can be overcome by a suitable choice of the X-ray source (fine focus X-ray tubes, $s < 1$ mm) or by increasing the distance from source to specimen (at the expense of the flux). In this case, and without taking into consideration the detector performance, diffraction will limit the resolution.

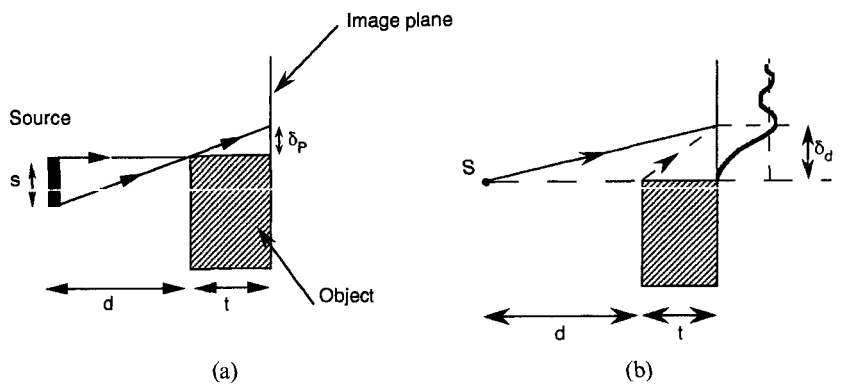


Figure 4. Factors affecting resolution in XCM. (a) Perumbral blurring; (b) diffraction effect.

Because of Fresnel diffraction, the specimen thickness (t) sets a fundamental limit to the resolution (Fig. 4b). In this case, the diffraction limit is given by

$$\delta_d = \sqrt{\lambda t} \tag{8}$$

where λ is the X-ray wavelength. No further improvement in performance can be made beyond this limit. Theoretical resolutions of a few tens of nanometers are obtainable in the soft X-ray range ($\lambda > 1$ nm) by using specimens of micrometer thickness.

2.2.3 Instrumentation

Basic XCM equipment requires only two devices: an X-ray source and a recording medium. Because the magnification is unity, the ultimate resolution is limited by the resolving capability of the detector, which is here the most important part of the system, the X-ray source having to fulfill the requirement of the detector.

2.2.3.1 Two-Dimensional Detectors

To approach the 'diffraction limited' performance with SXCM, high quality photosensitive recording media are necessary. Previously, this could not be done, not even with ultrafine-grain silver-halide emulsions (grain size $<1\,\mu$m). Ultimate resolution became attainable with the advent, in the late 1970s, of grainless films used in X-ray lithography for the research and development of very large scale integrated circuits (VLSI). These materials are high molecular weight polymer resists (photoresists), such as poly(methyl methacrylate), PMMA. Under X-ray irradiation, cross-linking or alteration of the polymer chain length occurs. The change in the degree of polymerization induces a change of the solubility. Thus chemical solvents allow the latent image to be revealed as a surface topography. For positive resists, the dissolution rate increases with the absorbed energy density per unit volume; for negative resists, the situation is reversed and unexposed areas are removed by etching [17, 18].

In a resist detector, the spatial resolution limit is determined by the energy and the range of the photoelectrons and Auger electrons which result from core level photoabsorption. In the soft X-ray range for which resists are efficient ($E < 1$ keV) such processes bring about a transfer of energy on a few nanometers range [19, 20], which is of the same order of magnitude as the diffraction limit for these wavelengths. So, one can expect a lateral resolution of the same order. Actually, the best achieved resolutions in SXCM are around 10 nm, which is very close to the theoretical limits imposed by the PMMA and Fresnel diffraction.

The high resolving power of the resists, if compared to photographic plates, must be paid for by loss of sensitivity. Roughly, for a given X-ray source, resists require 10^4 times the exposure of silver halide films. To avoid prohibitive exposure times necessarily implies the use of a high brightness X-ray source.

2.2.3.2 Resists: Processing and Readout

Positive resists, such as PMMA, the copolymer PMMA-MAA, or poly(butyl

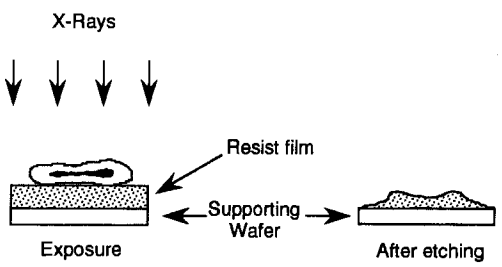


Figure 5. Principle of the resist method.

sulfone) (PBS), are commonly used for SXCM. When used, a uniform thin film of the resist (a few micrometers thick) is layered on a wafer (usually Si_3N_4, BN, or silicon) by a suitable technique.

As shown in Fig. 5, the specimen is placed in close contact to the resist. After exposure, the photoresist is carefully cleaned to remove the sample and developed in a solvent. This process must be monitored by differential interference microscopy (Nomarski) to achieve an optimal depth of development.

There are many ways to obtain subsequent magnification of the developed image. In particular, secondary electron microscopy (SEM), transmission electron microscopy (TEM), and atomic force microscopy (AFM) are mentioned here.

In principle, SEM observation requires the resist to be coated with a thin metal layer (gold or gold-palladium, for example) to avoid charging effects. But this procedure is irreversible, which means that if the resist was not in its optimum stage of development, no further process could take place. Moreover, electron-beam damage can occur in the resist for high acceleration voltage. To avoid these drawbacks SEM examination in the low-loss electron mode is more convenient. This mode does not require sputter coating; moreover, it enhances the contrast given by the small height features of the resist [21, 22].

A better resolution is to be expected from TEM observation although it requires a very thin support for the resist (carbon, Si_3N_4) or, alternatively, to float off the resist layer from the wafer. To avoid any artifact resulting from mechanical stress suffered by the resist during these manipulations and to preclude the inconveniences mentioned above concerning electron microscopy and especially damage caused by the beam during observation, techniques of surface replica have been developed [23–25].

AFM is the most recent technique used for resist examination [26]. It allows direct examination of the resist with no more preparation (i.e., no coating or replica). Moreover, it gives a high precision measure of the relief height which is the only reliable information for measuring the absorbed X-ray dose. These advantages will undoubtedly bring AFM into wider use in SXCM.

2.2.3.3 X-Ray Sources

Fluxes and spectral contents are two characteristic features of X-ray sources with which we are concerned. Cost and availability must also be taken into account. The main types of X-ray source used in XCM are 'synchrotron' radiation and 'plasma' sources.

Synchrotron radiation (SR) is an electromagnetic radiation generated by an electron beam moving on a curved closed trajectory in a synchrotron. For X-ray applications, special instruments devoted to intense SR emission have been developed. SR produces a broad-band

continuous spectrum; insertion of devices, such as undulators, allow line spectra to be produced which give higher X-ray fluxes at selected wavelengths. A complete description of SR properties can be found in the *Handbook on Synchrotron Radiation* [27].

SR is among the brightest X-ray sources available with typically 10^{12}–10^{14} photons $s^{-1}\,mm^{-2}\,mrad^{-2}$ (0.1% bandwidth). Nevertheless, literature on the subject (for instance [28]) shows that, even with such high fluxes, the required exposure times are relatively long. With monochromatized beams it takes several minutes. With broadband spectra (i.e., only low-pass filtering), these times are reduced to a split second, but at the expense of the contrast.

The relatively long exposure times required with SR is a major drawback, particularly for imaging the specimen during dynamic processes or for radiation-sensitive samples. Very slight modifications in the sample structures or small displacements during the exposure blur the image. This is most often the case for biological specimens. Living cells can move or, due to the bad yield of the resists, the specimens must suffer high radiation doses to be imaged. Thus, chemical changes and structural damages may occur in a very short time (<1 ns) [29].

Plasma sources offer an alternative to SR. Dense, hot plasmas are bright soft X-ray sources. Hot plasmas emitting in the X-ray range are attainable by means of high power pulsed lasers focused on a suitable target [30]. The brightness of an X-ray plasma source is typically 10^{13} to 10^{15} photons $pulse^{-1}\,mrad^{-2}$ in 1% bandwidth, which is much higher than standard synchrotrons. Under such conditions, the X-ray intensity emitted by the plasma is sufficient to fully expose a photoresist in one pulse (i.e., in a few nanoseconds). This ensures that the disadvantages of long exposure times are avoided [31–33].

2.2.3.4 Limitations of the Resist Method

XCM using resists is a very convenient way to obtain high resolution pictures but its lack of sensitivity is a notable disadvantage with consequences for the source requirement and radiation doses. The efficiency of the source and its restriction to the soft X-ray range makes the method unusable in many situations.

There are obviously other major inconveniences inherent in this technique. Thus, when using resists, multiple recordings of the same specimen are not easy, as the integrity of the specimen is not guaranteed owing to radiation damage or to mechanical stresses during the manipulation of the specimen. Multiple recording would be useful in the examination of dynamic processes or for elemental identification: multiple exposures at different X-ray energies permit the contrast on different atomic number species to be enhanced or allow element mapping, by subtracting two images recorded at energies below and above an absorption edge (dichromoradiography).

Another major disadvantage is the difficulty in deducing quantitative measurements from the contrast in the image: first, this contrast is dependent on the depth of development and, second, the response of the photoresists is nonlinear.

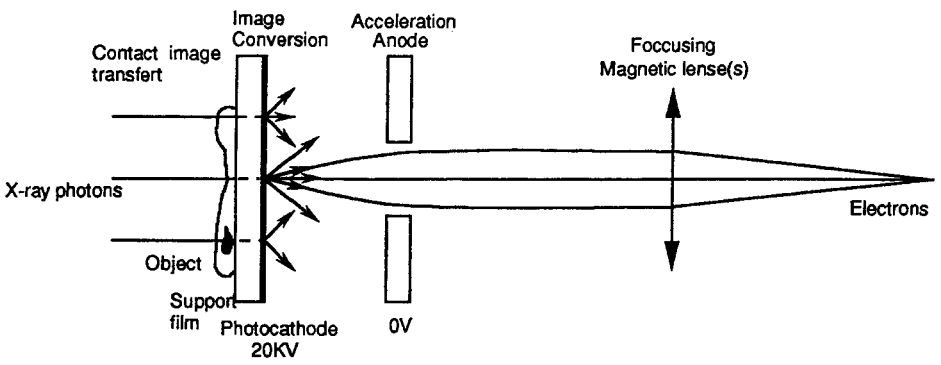


Figure 6. Principle of the contact photoelectron microscope.

2.2.3.5 Contact Photoelectron Microscope

To avoid some of the inconveniences encountered in the resist method, Polack and Lowenthal [34] proposed that the resist should be replaced by a photocathode layer. This has been implemented with a contact photoelectron X-ray microscope (CPEM). The experimental arrangement is depicted in Fig. 6. This consists of a plastic film which, on one side, supports the specimen and, on the other side, is coated with a thin photoemissive layer.

A convenient X-ray source casts a shadow X-ray image of the object on the photocathode. The photoelectron image is then enlarged by an electron optical system [35].

On the same basic principle, Kinoshita et al. [36] developed an X-ray detector called an *'X-ray zooming tube'*, which combines a CsI photocathode and an electron magnifying lens. After passing through two microchannel plates acting as image intensifiers, the electron image is converted into a visible light image by a fluorescent screen. Experiments on this apparatus with SR as X-ray source ($\lambda = 1.5$–15 nm) have been performed by Watanabe et al. [37]. The resolution was estimated to be around 0.2 μm. A commercial version of this prototype (Hammamatsu Photonics KK) is now available.

2.2.4 Applications of X-Ray Contact Microradiography

Most applications of XCM concern the biological sciences; applications of X-ray contact microscopy in materials science are still quite rare.

As seen above, the nonlinearity of the resists together with the difficulties in controlling the development process make the resist technique unsuitable for quantitative imaging, which considerably limits its application to materials science. However, quite opposite to the resists, and in spite of their nonlinearity, photographic plates as recording media with XCM have been largely used previously for elemental mapping and quantitative measurements. By using successively two excitation energies, one just above and the other just below a characteristic threshold, and by processing the two images, elemental maps can be obtained.

An application of elemental mapping has been given by Bigler et al. [38] with SR as X-ray source and polymetallic nodule samples. Kodak H-R plates were used as a detector. Figure 7 shows the maps (in negative) of Mn, Fe, and Ni obtained with a 30 μm thick section of such a nodule. The spatial resolution was $5 \times 5\,\mu m^2$ in a large field (up to a few cm^2). The elemental concentrations have been deduced from these results with an accuracy of a few percent. The sensitivity was estimated to be about 1 wt.%.

Another reason why the resist method is seldom used in materials science can be found in the fact that photoresists are efficient in the soft X-ray range and, consequently, this technique is restricted to the examination of thin, low density specimens. Thus Azuma et al. [39] have applied SXCM to the observation of grain boundaries, holes, and microspheres (selectively stained with osmium) in polymer composites.

The *biological sciences* are the most important field of application for SXCM. One of the first disciplines that took advantage of the technique was histology. Development and refinement of microradiography in histology has, to a large extent, been provided by Engström and Lindström [40, 11]. They have opened the way to soft X-ray microscopy and have laid the bases of quantitative cytochemical analysis by X-ray absorption spectroscopy.

Today, many authors emphasize the potential advantages of soft X-ray microscopies in general and of contact microscopy in particular: in the so-called 'water window' wavelength range (2.3–4.4 nm) X-rays give high-contrasted images on *unstained and unfixed* hydrated specimens.

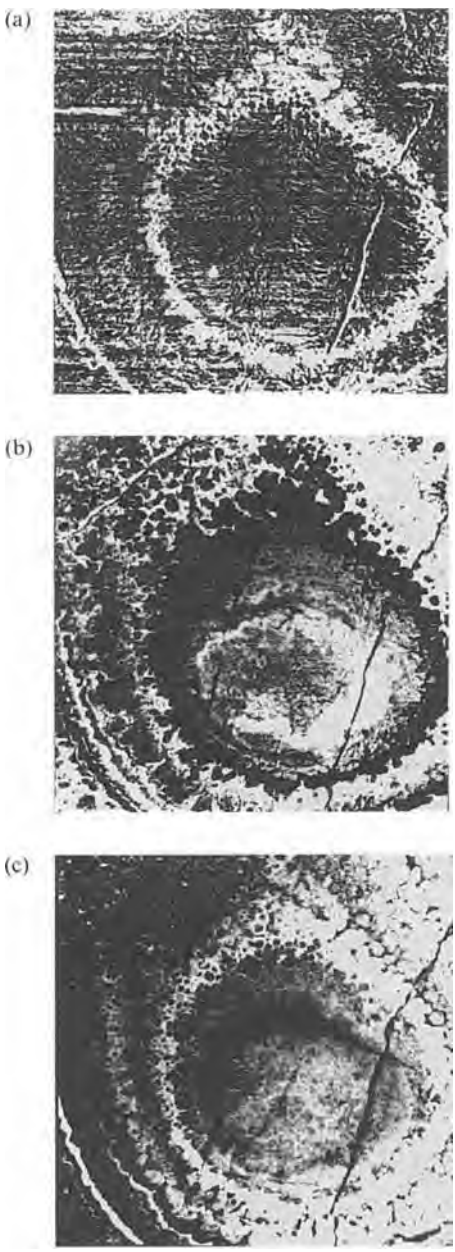


Figure 7. Characteristic images of a polymetallic nodule. Each elemental image is obtained after processing of two contact images taken at energies situated on either side of the K edge of the element. (a) Image showing concentration zones of Mn (in white). (b) Iron image. (c) Nickel image. Adapted from Bigler et al. [38].

At present, most of the applications of XCM concern imaging of living cells in the soft X-ray range with laser plasma sources. Interested readers will find the review by Ford et al. [41] to contain many such applications of XCM.

2.2.5 Conclusion

Contact microradiography is now becoming a widely used tool for qualitative imaging. Despite its great simplicity, XCM is the most resolving X-ray imaging technique since in some applications it can attain the theoretical limit of resolution set by fundamental constraints (diffraction and photoelectron range). The weak point of this technique is that quantitative measurement, only possible with photographic plates, cannot be reconciled with high resolution imaging. It is certain that advances in contact photoelectron microscopy will satisfy this requirement in the near future.

2.3 X-Ray Projection Microscopy

2.3.1 Principle

The point projection method (XPM), first suggested by Sievert [42], was implemented and developed for X-ray shadow imaging by Cosslett and Nixon [43]. With this method an enlarged radiograph is obtained by means of a point source of X-rays casting a shadow image of a nearby object onto a recording plane, such as a photographic film (Fig. 8). As for XCM, a further processing of the image (photographic or electronic enlargement) allows higher magnification to be obtained.

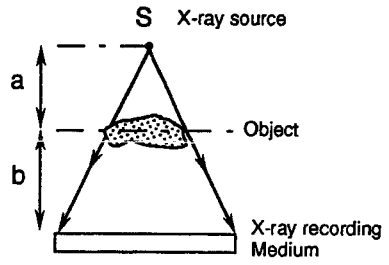


Figure 8. Principle of X-ray projection microscopy.

Originally the basic design of the instrument comprised a two stage magnetic electron lens to focus an electron beam on a metallic foil target a few micrometers thick. The X-rays generated by the electron impact on the target act as a high brightness microfocus X-ray source.

The technique became very popular in the 1960s and several instruments of this type were then commercialized (General Electric, TPD Delft, Microray, and others). At present a modernized version of basically the same apparatus is being developed.

2.3.2 Theoretical Considerations

2.3.2.1 Magnification and Resolution

The primary magnification factor is calculated from simple geometrical inferences. Using the notation given in Fig. 8, the magnification, M, is

$$M = 1 + b/a \tag{9}$$

Limiting factors to the lateral resolution are due to (i) the geometry and (ii)

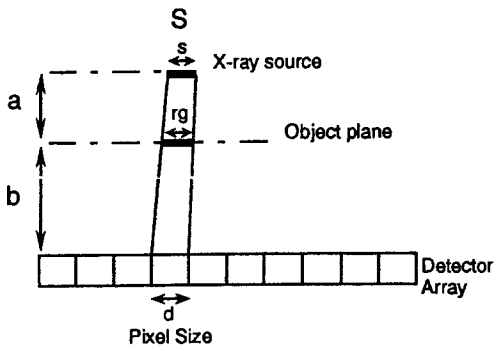


Figure 9. Geometrical contribution to the resolution in XPM (extended X-ray source and detector array).

Fresnel diffraction. These are described below.

The geometrical contribution to the resolution depends not only on the source size, s, but also on the picture element size, d (pixel). If we consider that the finest resolved features in the object cast in the image plane a total shadow whose size is d (Fig. 9), we can write

$$r_g = \frac{d}{M} + \left(1 - \frac{1}{M}\right)s \tag{10}$$

This clearly shows that, for high magnification projection geometry ($M \gg 1$), the *ultimate* resolution is given by the source size, s. In the limiting case of a point source ($s = 0$), the resolution would be limited by the pixel size ($r_g = d/M$).

Projected in the object space, the contribution of the Fresnel diffraction to the resolution limit is given by

$$r_d \simeq \sqrt{\frac{ab\lambda}{a+b}} \tag{11}$$

The *depth of field* is very wide (except for soft X-rays) but for other reasons, the specimen thickness must be limited. Most important is the transparency of the specimen to the radiation being used; also, as XPM gives a 2D representation of a 3D object, the depth information is lost and specimens which are too thick result in confused images not easily interpretable. This difficulty is worsened when the specimen is close to the source because the magnification varies greatly from one projected plane of the object to another. The only way to overcome these difficulties is to use a technique of reconstruction such as computed tomography (see Chap. 2, Sec. 3 in this Volume).

2.3.3 Instrumentation

Every X-ray projection microradiographic system is basically made up of a microfocus X-ray source, a sample holder attached to an X-Y-Z mechanical stage and an X-ray sensitive imaging detector. Because the lateral extent of the source must be reduced to the size of the smallest resolved detail in the object, XPM requires small-size X-ray sources. Microfocus X-ray sources, based on an electron beam focused on a metallic target, are used for that purpose.

Microfocus X-ray tubes are commercially available (e.g., Andrex, Feinfocus, and Philips). The high voltage typically ranges from 10 to 200 kV, and the maximum beam currents are in the region of a few milliamps. The focal spot size is adjustable with a minimum value of a few micrometers. The main advantage of such devices is the high X-ray power output they provide, which allows real-time microradiography. However, owing to the sealed window, the minimum distance of focus allowed to the object (~1 cm) limits in practice the magnification to about 100.

On the other hand, the energy range is restricted to hard X-rays, which prevents the use of such X-ray sources for the examination of low absorbing specimens. Moreover, X-ray tubes are polychromatic sources having characteristic lines superimposed on high *bremsstrahlung* (high Z target); the target is often not removable. This is why such sources cannot be used for elemental mapping or quantitative measurements that require the use of numerous selected X-ray characteristic lines.

The basic arrangement of Cosslett and Nixon (i.e., electron-gun and magnetic lenses) offers an alternative to commercial X-ray tubes. Such apparatus was commercialized in the 1960s. Today, these instruments are no longer available. Based on the same basic scheme, however, the use of a commercial scanning electron microscope (SEM) for projection X-ray microradiography, first suggested by Horn and Waltinger [44, 45], has attracted followers. With slight modifications of the sample holder, the electron beam in an SEM can be focused on a metallic foil sample. Acting as an anode, it constitutes a high brightness microfocus source. The X-rays generated are transmitted through the foil, the specimen being placed behind this foil in a vacuum, or, by using a suitable window, in an air environment.

The main advantages of the SEM with respect to X-ray tubes are as follows:

(i) It allows an easy control of the focus size; the optimal choice of the electron beam size must result from a trade-off between X-ray flux and resolution.
(ii) Easy lateral shifting of the X-ray source is possible. This proved to be very useful for stereo-microradiography.
(iii) It enables an easy change of the target material, which is essential for elemental mapping as well as for quantitative analysis purposes. In some instruments this facility is not available, the target being used as a vacuum separation window.
(iv) The sample holder can be a simple manual X-Y-Z stage, or equipped with computer controlled stepping motors. In this last configuration a precision rotation stage can be added, which makes microtomography possible.
(v) The detector may be a simple photographic camera or an X-ray sensitive electronic camera.

With regard to (v) a photographic film combines a large density of information and a large field of view. One disadvantage, however, is the long delay in processing. Even if this inconvenience is considerably reduced with the Polaroid instant system, the information medium is still not well suited to easy mathematical processing such as image difference, filtering, logarithmic compression etc.

Also with regard to (v) the X-ray sensitive electronic camera has tended to replace X-ray films. Image capture by the Newicon camera, coupled with an image intensifier, was performed by Newberry [46] using an old General Electric projection shadow microscope. Thomas et al. [47] used a phosphor screen coupled by a fiber optic faceplate with a low noise electronic camera to implement XPM in a commercial SEM. Their experimental arrangement is sketched in Fig. 10. A target holder carrying a set of different metallic foils allows the primary radiation to be easily changed. The imaging detector

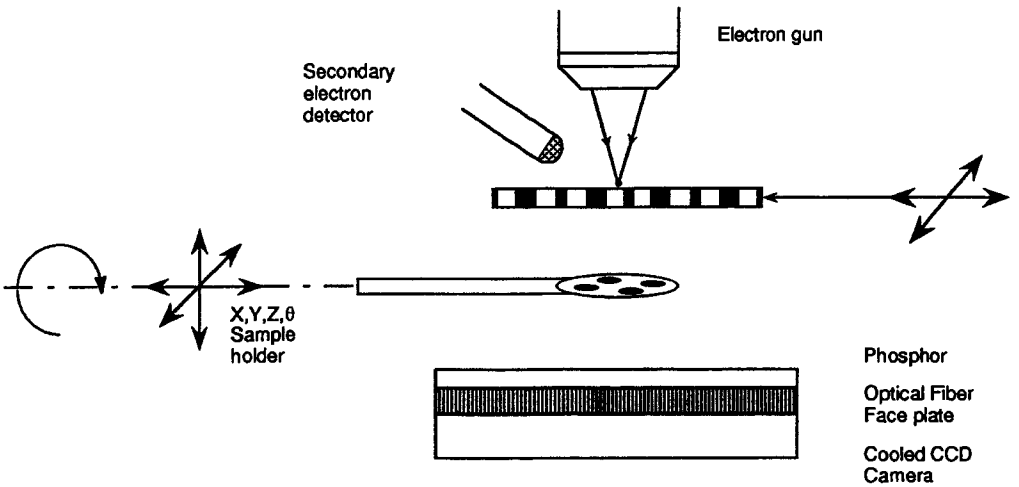


Figure 10. Image capture by electronic camera in XPM, after Thomas et al. [47].

is a CCD cooled camera, which permits acquisition times from a few seconds to tens of minutes, depending on the available X-ray flux. The specimen is kept under vacuum or in a special environmental chamber. A θ-rotation axis has been added to the standard 3D mechanical movement to perform microtomography.

The high linearity of CCD enables us to correct the data for dark current offset or fixed pattern noise resulting from non-uniformities of the response over the image field. So, a quantitative μt map of a specimen is easily obtained from the combination of the absorption image, the dark current image and the flat field image (taken without the specimen) [48].

Although based on the same principle, Cheng and coworkers [49] used a slightly different arrangement: the target acting as a window to maintain the vacuum inside the SEM chamber, and the specimen being outside in an air environment. The camera is a slow scan CCD coupled to a high resolution phosphor screen by a lens. The apparatus is devoted to three-dimensional imaging and can display stereo-pairs in real time or true tomographic reconstructions. Real time stereo images are skilfully obtained by oscillating the electron spot between two fixed points on the target; the two images, synchronously captured by a frame grabber, are successively displayed on a monitor. Special eyeglasses give a live stereo/video view of the object.

2.3.4 Discussion

2.3.4.1 Resolution

The resolution in XPM depends on (i) the size of the source and (ii) on the X-ray detector system.

The most important contribution to the X-ray source size comes from the spread of the X-ray generating range due to electron scattering in the target material. This corresponds to a lateral extent of a few

micrometers typically. In practice such spot sizes are obtainable with commercial X-ray tubes. With laboratory equipment smaller spot sizes can be obtained. For that purpose high-Z target elements can be used with a lower electron beam energy or, alternatively, thin film layers deposited on a low-Z substrate can be used [49–51]. In both cases, but more particularly in the latter, the resolution is improved at the expense of the X-ray flux. To avoid impracticable lengthy acquisition times a balance between flux and resolution has to be found so that the ultimate resolution attainable stays at about 0.2 μm.

In photographic recording, owing to the smallness of the grain size, the major contribution to the resolution is the source size. This gives the ultimate resolving power. With a CCD camera, the pixel size, the optical coupling system, and the phosphor screen are deciding factors in the final resolution. The actual performance of the whole system can be characterized by its modulation transfer function (MTF). For phosphor screens, the MTF is roughly inversely proportional to the phosphor thickness. To reduce this thickness implies a lower conversion efficiency, so a compromise must be found. Typically, the thicknesses being used vary from 20 to 50 μm, depending on the energy of the X-ray used and MTF of about 20 line pairs per millimeter are measured (5% modulation). Moreover, pixel sizes are typically 20 μm × 20 μm. Thus the resolution of X-ray cameras (in the image plane) stays about 50 to 100 μm. Bearing in mind Eq. (10), we can estimate that, to obtain a resolution of, say, 1 μm, the geometrical enlargement must be 50 or 100. Owing to the limited distance between the focal spot and the camera (a few centimeters to have a sufficient flux) we see that the distance from the target to the object must be reduced to a few hundred micrometers. Under these conditions it is clear that sub-micrometer resolutions are not so easy to obtain with electronic cameras.

Other types of area X-ray detectors based on nonelectronic systems are under investigation or have already been developed. Among them, imaging plates (Fuji, Kodak), which are erasable systems based on photostimulated luminescence, present a great interest (high linearity, high dynamic range), and they are likely to become widely used in industrial and medical radiology. Their limited spatial resolution (~100 μm) makes them less interesting for XPM at the present time.

2.3.4.2 Microanalysis

Quantitative X-ray microanalysis by XPM has been widely applied, particularly in biological fields. Elemental X-ray mapping is currently obtained by the differential technique [11–13]. However, *true analytical* X-ray microscopy, which combines imaging with multielemental analysis, is still a challenge. A new approach that allows quantitative elemental mapping and, independently, thickness mapping has been recently proposed and experiments are in progress [16, 52].

2.3.5 Applications

High energy microfocus radiography (20–200 keV) is widely used in industrial or medical applications. For these

applications, a commercial instrument (e.g., Andrex, Hamamatsu, or Philips), which includes a conventional microfocus X-ray source (spot size $<10\,\mu m$) and an X-ray video camera, allows a real-time X-ray image of the object with a resolution of a few tens of micrometers to be obtained.

Industrial applications concern the control of products (e.g., castings, ceramics, electronic components) to detect flaws such as microporosites, cracks or defective solder joints in parts. In materials science research, microradiography has been used for the study of alloys, composite materials etc. XPM has also been applied to the study of dynamical phenomena. Images of electrochemical processes, such as metallic deposition during electrolytic plating, have been obtained [53]. In the same way, dynamic studies of metal solidification have been performed [54].

In the medical sciences the applications of microradiography are numerous: cancer lesion studies, evaluation of bone mineral density in the pathology of osteoporose by dual energy X-ray absorptiometry (DEXA), inspection of vascularized tissues by microangiography, teeth studies etc.

At *medium* and *low energies* (1–20 keV) a Cosslett and Nixon (prototype) arrangement must be used. Most of the proposed applications are within the materials science domain (e.g., alloy composition studies, cast iron, composite materials etc., Fig. 11), geological sciences (e.g., microfossil radiographs), and biological applications (e.g., microradiographs of insects, plants etc.). Dynamical studies have also been performed: the intercalation process of acids in graphite [48], solid–solid diffusion [55], and ion diffusion [56].

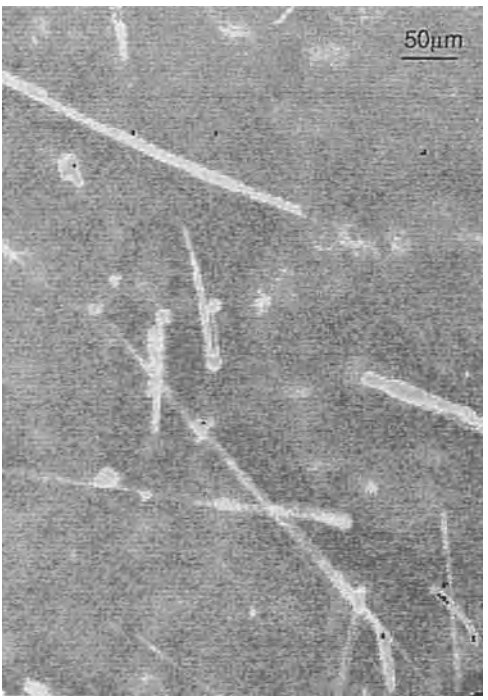


Figure 11. Zinc precipitates in a magnesium matrix obtained with an electronic camera in XPM (see Fig. 10). (Photo: Courtesy of H. Elhila.)

2.3.6 Conclusion

X-ray projection microscopy, in spite of its poor resolving capability, presents many advantages when compared with X-ray contact microscopy. The following conclusions are worth noting:

- Projection microscopy, unlike contact microradiography, offers a large depth of focus (except for long wavelengths as seen before) so that thicker samples can be examined.
- The *noise* in the recorded images, due to scattered radiation, diffraction, fluorescence, or originating from a radioactive element in the sample, is considerably reduced in XPM because these radiations are collected by the whole surface

of the detector over a relative small solid angle. In XCM, unlike XPM, all these types of radiations are collected in the half-space by a minute surface of the detector. Consequently, the contrast in X-ray projection images is better and the image more easily interpreted.

- XPM presents many practical advantages. For example, the magnification can be changed at will by a simple displacement of the specimen along the optical axis. With X-ray sensitive cameras, electronic zooming is possible; successive images can be registered to follow a dynamic process. *Stereoradiography* and *dichromoradiography* (i.e., two images taken at different energies) are easy to implement and this is of great importance in quantitative analysis.
- A major advantage of XPM, using electronic cameras, is that microradiography and microtomography can be implemented in the same apparatus. Taking into account the potential analytical capabilities of this technique, XPM is emerging as a powerful method of investigation.

2.4 References

[1] G. J. Burch, *Nature* **1896,** *122*, 841.
[2] F. Ranwez, *C. R. Acad. Sci. Paris* **1896,** *122*, 841.
[3] B. K. Agarwal, *X-Ray Spectroscopy*, Springer Verlag, Berlin **1979**.
[4] W. M. J. Veigele, *At. Data* **1973,** *5*, 52.
[5] B. L. Henke, P. Lee, T. J. Tanaka, R. L. Shimabukuro, B. K. Fujikawa, *At. Data Nucl. Data Tables* **1982,** *27*, 1.
[6] J. H. Hubbell, *NSRDS – NBS circular 29*, US Department of Commerce **1969**.
[7] E. Storm, H. I. Isreal, *Nucl. Data Tables A 7*, Academic Press, New York **1970,** p. 565.
[8] B. L. Henke, A. Gullikson, B. Davis, *At. Data Nucl. Data Tables* **1993,** *54*, 181.
[9] T. P. Thinh, J. Leroux, *X-Ray Spectrom.* **1979,** *8*, 85.
[10] K. F. J. Heinrich in *Proc. 11th Int. Congress on X-Ray Optics and Microanalysis* (Eds. J. D. Brown, R. H. Packwood) **1986,** 67.
[11] B. Lindström in *X-Ray Optics and X-Ray Microanalysis* (Eds. M. M. Pattee, V. E. Cosslett, A. Engström), Academic Press, New York **1963,** p. 13.
[12] C. J. Buckley, G. F. Foster, R. E. Burge, S. Y. Ali, C. A. Scotchford in *X-Ray Microscopy III* (Eds. A. G. Michette, G. R. Morrisson, C. J. Buckley), *Springer Series in Optical Sciences*, Vol. 67, Springer Verlag, Berlin **1992,** 423.
[13] E. Bigler, F. Polack, *J. Phys. C Suppl.* **1984,** *45*, 83.
[14] S. Doniach, P. Eisenberger, K. O. Hodgson in *Synchrotron Radiation Research* (Eds. H. Winick, S. Doniach), Plenum Press, London **1980,** pp. 223–275.
[15] L. Grodzins, *Nucl. Instrum. Methods* **1983,** *206*, 541.
[16] J. Cazaux, *Microsc. Microanal. Microstruct.* **1993,** *4*, 513.
[17] E. Spiller, D. E. Eastman, R. Feder, W. D. Grobman, W. Gudat, J. Topalian, *J. Appl. Phys.* **1976,** *47*, 5450.
[18] A. R. Neureuther in *Synchrotron Radiation Research* (Eds. H. Winick, S. Doniach), Plenum Press, London **1980,** pp. 223–275.
[19] R. Feder, E. Spiller, J. Topalian, *J. Polym. Eng. Sci.* **1977,** *17*, 385.
[20] E. Spiller, R. Feder in *X-Ray Optics, Applications to Solids* (Ed. H. J. Queisser), *Topics in Applied Physics, Vol. 22*, Springer Verlag, Berlin **1977,** pp. 35–92.
[21] P. C. Cheng, D. M. Shinozaki in *X-Ray Microscopy III* (Eds. A. G. Michette, G. R. Morrisson, C. J. Buckley), *Springer Series in Optical Sciences*, Vol. 67, Springer Verlag, Berlin **1992,** 359.
[22] O. C. Wells, P. C. Cheng in *X-Ray Microscopy II*, Springer Verlag, Berlin **1988,** pp. 316–318.
[23] J. Kirz, D. Sayre in *Synchrotron Radiation Research* (Eds. H. Winick, S. Doniach), Plenum Press **1980,** pp. 277–322.
[24] P. C. Cheng in *X-Ray Microscopy – Instrumentation and Biological Applications* (Eds. P. C. Cheng, G. J. Jan), Springer Verlag, Berlin **1987,** pp. 65–104.
[25] K. Shinohara, S. Aoki, M. Yanagihara, A. Yagishita, U. Iguchi, A. Tanocka, *Photochem. Photobiol.* **1986,** *44*, 401.
[26] T. Tomie, H. Shimizu, T. Majima, M. Yuamada, T. Kanayama, H. Kondo, M. Yano, M. Ono, *Science* **1991,** *252*, 691.

[27] E. E. Koch, D. E. Eastman, Y. Farge in *Handbook on Synchrotron Radiation*, Vol. 1A (Ed. E. E. Koch), North-Holland, Amsterdam **1983,** pp. 1–59.
[28] K. Shinohara, *Proc. SPIE* **1989,** *1140*, 196.
[29] R. A. London, M. D. Rosen, J. E. Trebes, *Appl. Opt.* **1989,** *28*, 3397.
[30] N. M. Ceglio, *Ann. N.Y. Acad. Sci.* **1980,** *342*, 65.
[31] T. W. Ford in *X-Ray Microscopy III* (Eds. A. G. Michette, G. R. Morrisson, C. J. Buckley), *Springer Series in Optical Sciences*, Vol. 67, Springer Verlag, Berlin **1992,** pp. 438–441.
[32] M. S. Shulz, A. G. Michette, R. E. Burge in *X-Ray Microscopy III* (Eds. D. Sayre, M. Howell, J. Kirz, H. Rarback), *Springer Series in Optical Sciences,* Springer Verlag, Berlin **1992,** pp. 58–61.
[33] H. Kondo, T. Tomie, *J. Appl. Phys.* **1994,** *75*, 3798.
[34] F. Polack, S. Lowenthal in *X-Ray Microscopy* (Eds. G. Schmahl, D. Rudolph), *Springer Series in Optical Sciences,* Vol. 43, Springer Verlag, Berlin **1984,** 251.
[35] F. Polack, S. Lowenthal, D. Phalippou, P. Fournet in *X-Ray Microscopy II* (Eds. D. Sayre, M. Howell, J. Kirz, H. Rarback), *Springer Series in Optical Sciences*, Vol. 56, Springer Verlag, Berlin **1988,** 220.
[36] K. Kinoshita, T. Matsumura, Y. Hirai, M. Sugiyama, H. Kihara, N. Watanabe, Y. Shimanaki, A. Yagashita in *X-Ray Microscopy III* (Eds. D. Sayre, M. Howell, J. Kirz, H. Rarback), *Springer Series in Optical Sciences,* Vol. 67, Springer Verlag, Berlin **1992,** pp. 335–337.
[37] N. Watanabe, T. Matsumura, Y. Inagaki, K. Kinoshita, Y. Shimanuki, K. Furuya, T. Taguchi, M. Taniguchi, H. Kihara, *J. Microsc.* **1993,** *170*, 141.
[38] E. Bigler, F. Polack, S. Lowenthal, *Opt. Commun.* **1982,** *41(1)*, 6.
[39] H. Azuma, A. Takeichi, S. Noda, *Jpn. J. Appl. Phys.* **1994,** *33*, 4622.
[40] A. Engström, *Acta Radiol., Suppl.* **1946,** *63*.
[41] T. W. Ford, A. D. Stead, R. A. Cotton, *Electron Microsc. Rev.* **1991,** *4*, 269.
[42] R. Sievert, *Acta Radiol.* **1936,** *17*, 299.
[43] V. E. Cosslett, W. C. Nixon, *J. Appl. Phys.* **1953,** *24*, 616.
[44] H. R. F. Horn, H. G. Waltinger, *Rem. Bedo.* **1973,** *6*, 163.
[45] H. R. F. Horn, H. G. Waltinger, *Scanning* **1978,** *1*, 100.
[46] S. P. Newberry in *X-Ray Microscopy III* (Eds. D. Sayre, M. Howell, J. Kirz, H. Rarback), *Springer Series in Optical Sciences,* Vol. 67, Springer Verlag, Berlin **1992,** pp. 306–309.
[47] X. Thomas, J. Cazaux, D. Erre, D. Mouze, P. Collard in *X-Ray Microscopy III* (Eds. D. Sayre, M. Howell, J. Kirz, H. Rarback), *Springer Series in Optical Sciences,* Vol. 67, Springer Verlag, Berlin **1992,** pp. 190–194.
[48] D. Erre, D. Mouze, X. Thomas, J. Cazaux, *Inst. Phys. Conf. Ser.* **1992,** *130*, Ch. 7, p. 567.
[49] P. C. Cheng, in *X-Ray Microscopy III* (Eds. D. Sayre, M. Howell, J. Kirz, H. Rarback), *Springer Series in Optical Sciences,* Vol. 67, Springer Verlag, Berlin **1992,** pp. 184–189.
[50] K. Yada, S. Takahashi, *J. Electron Microsc.* **1989,** *38*, 321.
[51] V. Kolarik, V. Svoboda, *J. Microsc.* **1989,** *156*, 247.
[52] P. Trebbia, J. M. Wulveryck, N. Bonnet, *Microbeam Anal.* **1995,** in press.
[53] M. Wevers, P. De Meester, M. Lodewijckx, Y. Ni, G. Marchal, Y. Jiang, J. Dequeker, P. Gensens, H. Vandeursen, D. De Ridder, L. Baert, G. Pittomvils, R. Boving, *NDT&E Int.* **1993,** *26(3)*, 135.
[54] W. F. Kaukler, F. Rosenberger, *Metall. Matter. Trans. A* **1994,** *25*, 1775.
[55] J. Cazaux, D. Erre, D. Mouze, J. M. Panat, S. Rondot, A. Sasov, P. Trebbia, A. Zolfaghari, *J. de Phys. IV*, Coll. C7, sup. J. de Phys. III, *vol. 3*, Nov. **1993,** 2099.
[56] S. Rondot, J. Cazaux, O. Aaboubi, J. P. Chopart, A. Olivier, *Science* **1994,** *263*, 1739.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

3 X-ray Microtomography

3.1 Introduction

In Sec. 2 of this Chapter it was explained that X-ray microscopy allows the study of the internal structure of an object. Unfortunately, X-ray images are two-dimensional projections and so fail to provide an adequate idea of this microstructure in three dimensions. Apart from the size of the object and the range of the lateral resolution, the same problem holds for X-ray radiography for which the need to see the interior of the specimen, without opening it, is obvious when this technique is applied in medicine. This problem was solved in the 1970s by the discovery of computed tomography (CT) [1], originally employed as a medical diagnostic tool [2], which is now being used increasingly in industry [3]. The goal of X-ray microtomography is exactly the same as for CT but with an improved resolution (i.e., the μm range instead of the 0.2 mm range) and for smaller objects (i.e., of the order of mm instead of the 10 cm range and using photon energies in the 5–30 keV range instead of energies above 80 keV). Consequently, X-ray microtomography may be considered as a combination of X-ray microscopy and CT. The initial signals result from the same photon–matter interaction as in XPM and XCM (see Fig. 1 of Sec. 2 of this Chapter), while the data acquisition and processing are inspired by CT.

The basic principles of CT are illustrated in Fig. 1 for the simple case of the reconstruction of a single transverse section of a specimen irradiated by a parallel and monochromatic X-ray beam. It is helpful to imagine the specimen subdivided into a matrix of pixels, $N \times N$, each of them being characterized by its absorption coefficient μ_{ij}, which has to be determined [4, 5]. When the source and the detector are in the position 1, the intensity measured by an element detector k, I_{1k}, results from the absorption of X-rays along their path through the N pixels of the column i (hatched area in Fig. 1), in such a way that Beer's law allows us to write

$$i_{1k} = \ln(I_0/I_{1k}) = \sum_{j=1}^{N} \mu_{ij} r \qquad (1)$$

where I_0 is the incident intensity and r is the common dimension of each pixel. Changing, step-by-step, the angular position of the coupled source and detector with respect to the object, the intensity measured by the same element detector k, results in a similar summation but involving other pixels of the object. For instance, in the position m, the intensity

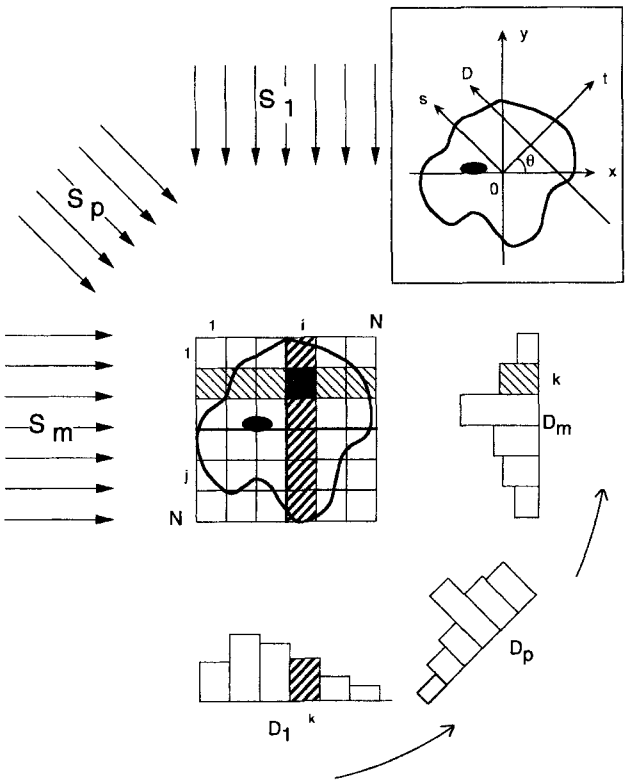


Figure 1. The principle of CT. A transverse section of the specimen is subdivided into $N \times N$ pixels. The source-(linear) detector pair is rotated around the specimen from the position S_1D_1 to S_pD_p, S_m, D_m,... The acquisition of the corresponding absorption profiles allows a determination of the linear absorption coefficient μ_{ij} of each pixel. The Radon transform (insert) relates the coordinates x, y of a point of the object to the oblique path, SD, of an X-ray beam.

I_{mk} is given by

$$i_{mk} = \ln(I_0/I_{mk}) = \sum_{j=1}^{N} \mu_{ij} r \qquad (2)$$

A set of linear equations is then established correlating the i_{mk} values to the μ_{ij} values and, in principle, a matrix inversion allows us to deduce the unknown values μ_{ij} from the measured quantities i_{mk}. This is possible if the number of measurements is equal to the number of the unknown values, implying the use of an array of N detectors and N different angular positions using an angular step $\Delta\theta$ where $\Delta\theta = 180°/N$. To form images of a cross sectional slice of the specimen of 0.5 mm in diameter with a pixel size (or resolution) of $r \approx 5\,\mu$m, the acquisition of around one hundred profiles on a linear detection system composed of one hundred elementary detectors is required and the angular step is of the order of 2°.

The same philosophy is easily adapted to the situation where projection images (instead of profiles) are acquired by the use of a two-dimensional array of detectors to reconstruct the object in three dimensions by the determination of the μ_{ijk} value of each of its voxels (volume elements).

From these general principles, it can be deduced that there are two different aspects in X-ray microtomography: the instrumental details of the image acquisition and the mathematical details of the reconstruction of the object. These aspects are the subject of the following subsections although it must be kept in

mind that:

(i) The resolving power in microtomography is not only a function of the resolution of the initial images (2D detector) or profiles (1D detector) but also of the number N of different views being acquired: $r = t/N$ where t is the typical dimension of the region of the object to be reconstructed.

(ii) The sensitivity $\Delta(\mu t)/\mu t$ for a projection image is associated with a given dose D_0 (= n_0 incident photons per pixel of dimensions $r \times r$ in the object plane) (see Eq. (5), Sec. 2.1 of this Chapter). In three dimensions, i.e., for a voxel $r \times r \times r$, the same resolution requires a cumulated dose ND_0 but the statistical noise is then increased by $\sqrt{N}$ so that the sensitivity is deteriorated by the same factor. The result is that the dose has to be increased by $N^{1.5}$ to maintain the resolution r and the sensitivity constant when going from two to three dimensions. In fact instead of $\sqrt{N}$, the noise amplification of the reconstruction procedure is [6–9]:

$$\omega \approx B\sqrt{N} \tag{3}$$

where B is about 2 for the best reconstruction algorithms [5]. Then the relative error in the projection data of 1% requires $n_0 \approx 10^5$ photons per pixel and 2×10^8 photons are needed for $N \approx 100$ to achieve the same sensitivity in CT measurements.

In the same manner, the improvement of the lateral resolution by a factor k ($k < 1$) in a projected image requires an increase in the dose of $k^{-2}D_0$ to keep the number n_0 of incident photons per pixel and the sensitivity constant. In three dimensions the corresponding increase in dose becomes $Bk^{-3.5}D_0$. As can be seen from Fig. 2, the conclusion of this simple analysis is that any slight improvement of the resolution in CT has to be paid for by a large increase of the data acquisition time (for a constant photon flux on the specimen).

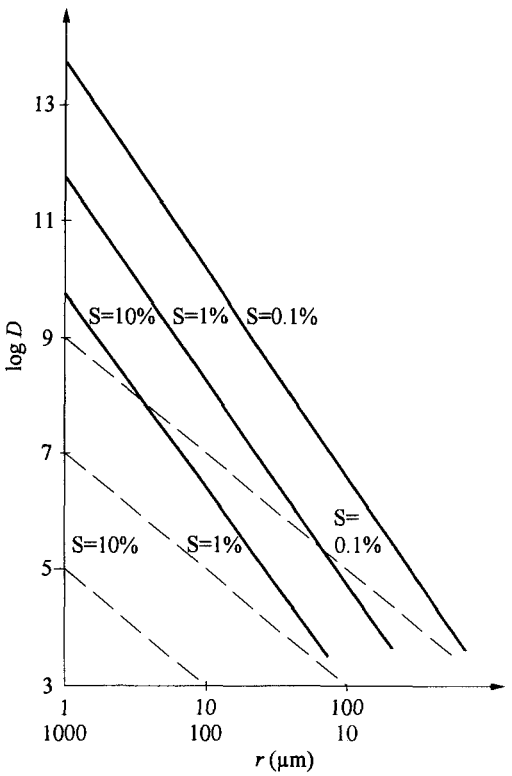


Figure 2. Cumulated dose (log scale and in photons μm^{-2}) required for reaching a given resolution r (μm) and a given sensitivity S (%) in X-ray microtomography (full lines) compared to that required for only one projection, in X-ray microscopy (dashed lines). The typical dimension of the specimen is supposed to be 1 mm.

(iii) Equations (1) and (2) only hold for the attenuation of a monochromatic X-ray beam. Then, it is necessary to deal with monochromatic projections to perform quantitative X-ray microtomography. If not, artefacts from different attenuations of different photon energies, like the beam hardening effect which is associated with

the lower attenuation of X-ray photons of higher energies.

Consequently, in microtomography the performance is mainly characterized by the resolving power, the dimensions of the volume to be reconstructed and the corresponding data acquisition time, with also the sensitivity being achieved and the possibility to perform quantitative 3D imaging.

3.2 Instrumental Details

The microtomography set-up, as with the commercially produced tomography systems for medical or industrial purposes, does not require the use of X-ray focusing elements. This set-up can be arbitrarily subdivided into three parts: the radiation source, the object scanning mechanism, and the detector system [10]. The source–detector pair moves around a horizontal immovable specimen in the serially manufactured medicinal scanners, while industrial CT systems and microscanners are mainly based on the rotation (and translation) of only the specimen.

When the source is synchrotron radiation, the monochromatic X-ray beam is nearly parallel and the lateral resolution of the initial images or profiles is governed by the dimension D of an elementary detector which is presently of around 3–5 μm [9]. When more classical microfocus X-ray sources are used, the geometry nearly corresponds to the fan (or pseudo fan) geometry of tomography systems. For these arrangements, and as for X-ray projection microscopy, the lateral resolution of the initial images is $r_g = s + (D - s)/M$ where s is the source size and the magnification coefficient, M, is close to 2 when the specimen is set at the mid-distance between the source and the detector. This resolution may be improved by approaching the specimen from the source (thus increasing M).

The detailed object movements are closely related to the choice of the detection system. Like in medical CT, this last choice allows us to define various generations for the instruments [11] which are illustrated in Fig. 3.

(i) The first generation is based on the use of a single photon counting detector. This solution requires the detector, or the specimen [12, 13], or the X-ray source [10] to be translated to obtain a single profile. The acquisition of each profile is then followed by step rotations of the specimen in order to reconstruct a single slice defined by the direction of translation and the X-ray source. This solution is obviously time consuming but it allows us to overcome the difficulty associated with the beam hardening effect when the detector is also used as the analyzer (as is the case with an EDS solid-state detector) or when the source is monochromatized [12].

(ii) The second generation is based on the use of a linear detector array of, for example, 256 photodiodes in front of which an X-ray–visible phosphor converter has been placed. In this situation just a stepped rotation of the specimen is needed to acquire the various profiles required to form the image of a cross-sectional slice parallel to a collimated X-ray fan beam.

For a 3D reconstruction of the specimen, the acquisition of each sectional slice has to be followed by a step translation normal to the corresponding plane.

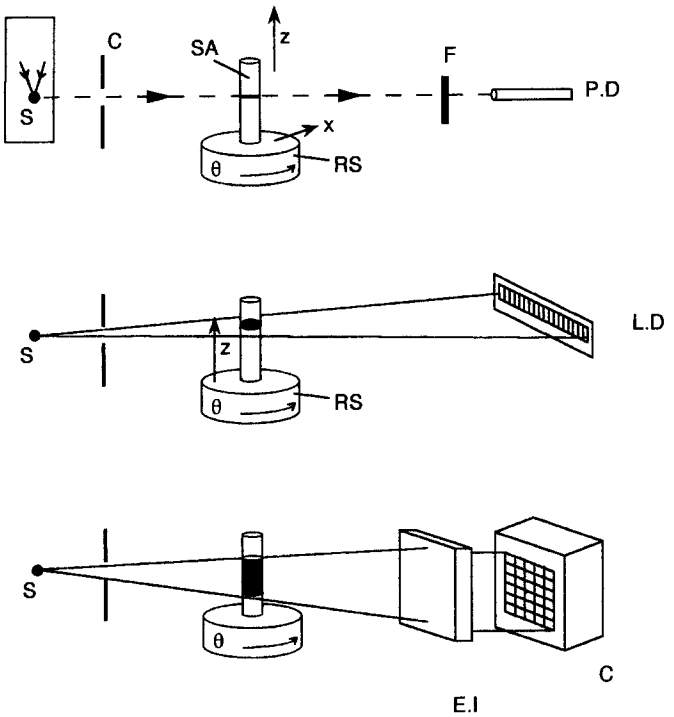


Figure 3. Architectures of the three generations of microscanners. S: X-ray source; C: collimator; RS: rotating stage; SA: sample; F: filter; PD: point detector; LD: linear detector; C: camera; EI: scintillator. The black regions of the sample correspond to those irradiated by the detected beam. For a 3D reconstruction, in addition to the common rotation (θ) the sample has to be translated along the x and z directions in the first generation (top), and only along the z direction in the second generation (middle).

(iii) The last generation of detectors is composed of a 2D matrix of individual elements which allows the acquisition of a shadow projection image of the specimen. Only a stepped rotation of the specimen is needed to acquire all the data for a next 3D reconstruction of the object.

The variety of architectures for the instruments presently in operation in the world results from the number of possible arrangements of the source and the detectors. For laboratory set-ups, there are, first, systems configured around conventional X-ray diffractometers [12] and microfocus X-ray sources (from Fe $K_\alpha = 6.4$ keV to Mo $K_\alpha = 17.45$ keV) combined with the use of a single photon detector [13] or a two-dimensional detector [14, 15]. Some other laboratory set-ups are derived from a microfocus X-ray radiography instrument and allow the acquisition of two-dimensional projections via the use of an image intensifier [16].

Taking into account the fact that the number of photons received per pixel in the image plane decreases as the inverse square of the distance from the source to the detector (when divergent X-ray sources are used), more compact instruments were highly desirable. Sasov [10] uses a scanning electron microscope (SEM) as a scannable X-ray source, setting the specimen inside the vacuum of the microscope. He also suggested various experimental arrangements based on the use of a single (or a linear array of) detector(s). The scanning of the electron beam along an inclined target allows us to change the view by only rotating the specimen, without any translation of the specimen or detector. Also derived from the architecture of an

SEM, the latest generation of laboratory microscanners are projection X-ray microscopes where the specimen can be rotated by a stepper motor and where the direct acquisition of digital absorption images is performed by the use of a CCD camera. The use of a thin foil target irradiated on one side by the incident electron beam, with the specimen set on the other side allows us to have an X-ray source in the micrometer range (see also Sec. 2.3 of this Chapter). Based on these general principles, two set-ups are presently under operation. In the first set-up, the foil target is also used to separate the SEM chamber from the normal atmosphere where the specimen is set [17]. In the second set-up, the specimen is set in the vacuum of the SEM chamber but the target foil can be easily changed and a projection image can be acquired in less than 10 s because of the short distance, a few centimeters, between the source and the detector [18].

Finally, there are microscanners based on the use of a monochromatic synchrotron radiation in which the parallel transmitted X-ray beam is collected via an array of detectors [19] or better via a CCD camera. Figure 4 illustrates the architecture of the arrangement used in this last situation [9, 20–24].

As expected from point (ii) of Sec. 3.1 of this Chapter (and Fig. 2), there is not a considerable difference in the lateral resolution of the reconstructed images obtained by these various solutions: a few micrometers for the latest sophisticated instruments based on the use of synchrotron radiation and of a CCD camera [22, 23] and a few tens of micrometers for the first generation of microscanners based on a microfocus X-ray

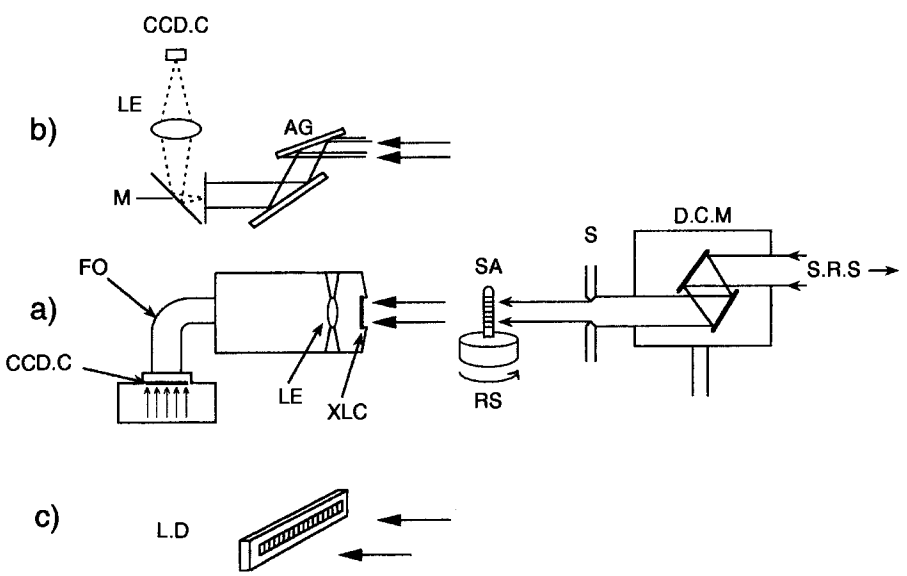


Figure 4. Set-up for microtomography using synchrotron radiation. (a) S.R.S: synchrotron radiation source; D.C.M: double crystal monochromator; S: sht; RS: rotating state; SA: sample; XLC: X-ray/light converter; LE: optical lens; FO: fiber optic bundle; CCD.C: camera (inspired from Kinney et al. [21]). (b) AG: asymmetrical grove crystal for X-ray optical enlargement; M: mirror (inspired from Bonse et al. [22]).

source and a single photon detector [25]. But in the first case the data acquisition time is restricted to a few minutes for a 3D reconstruction of the whole object compared to several hours for the restitution of a single cross section. If these remarks lead to the conclusion that the use of 2D detectors allowing the direct acquisition of projection digital images is more important than the use of the synchrotron radiation as an X-ray source, it must be kept in mind that SR sources can be tuned. The key advantage of X-ray microtomography based on SR sources is its potential to perform *quantitative* microtomography, sensitive to different elements (by utilization of absorption edges) and to different chemical phases (by using the X-ray absorption near edges structures) [26].

3.3 Overview of Image Reconstruction Methods

3.3.1 General Principles

Being independent from the real dimensions of pixels or voxels, the problem of image reconstruction from projections in microtomography is the same as in conventional CT for homothetic geometries. From the early works of Hounsfield [1], this remark explains the proliferation of papers related to this problem combined with the exploding capabilities of modern computers [3–5, 27–29].

Instead of the use of the matrix approach involving the discrete elements $\mu_{ij(k)}$ of Eq. (1), the general approach deals with the Radon transform associating the continuous change of the linear absorption coefficient μ expressed as a function of the coordinate system of the object x, y, (z) to the attenuation of the intensity of an oblique X-ray beam along its path length s characterized by the polar variables θ, t (see insert of Fig. 1). Then, in two dimensions, the equivalent of Eq. (1) is of the form [4, 5, 28, 29]

$$i(t,\theta) = \int_s \mu(x,y)\,\mathrm{d}s \tag{4}$$

and the image reconstruction problem is to determine $\mu(x,y)$ from $i(t,\theta)$ or, in other words, it is to find the inverse Radon transform of $i(t,\theta)$.

For this goal the existing algorithms can be categorized into two broad classes: analytic and algebraic [4, 5, 28, 29]. The back projection methods are the most popular of the analytic class, widely used in medical CT. The density profile in each direction is projected back along the linear path. These densities combine to produce an approximate map of the original structure. But such a procedure leads to artefacts resulting from the use of a finite number of projections N. For instance, for a set of isolated small objects with a small number of projections N, this method gives star patterns, each star being composed of $2N$ branches (Fig. 5). Such a result may be understood by the use of the projection slice theorem from which a projection taken at angle θ in real space corresponds to volume elements along a line (the slice) at the same angle in frequency space. The combined results from many directions do not completely fill Fourier space and miss out the high spatial frequencies. The result is the need to use filters to minimize these types of artefacts and the filtering operation can be performed either in real space or in the

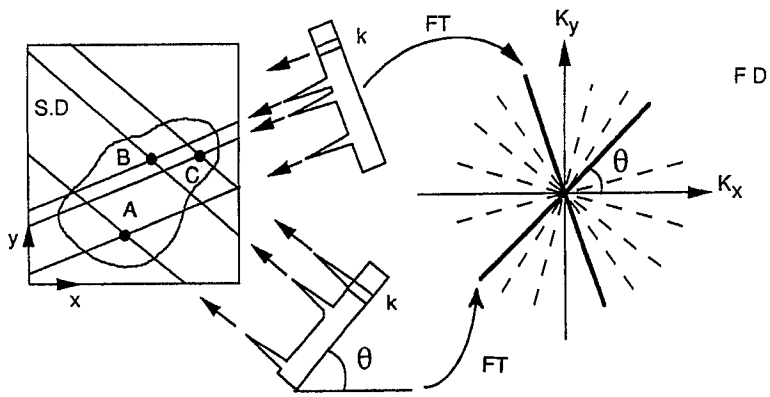


Figure 5. Back-projection schematic diagram. The specimen is composed of three strongly absorbing points (A, B, C) in the spatial domain SD on the left. Back-projecting the profiles acquired from N different θ directions produces a complete tomogram in SD. (For clarity only two profiles are shown in the Figure.) The initial point details look like stars at $2N$ branches in the tomogram. This artefact is easily understood when each projection is converted to its Fourier domain, where the finite number of projections leaves a gap in this frequency space. Filtering operations have to be implemented.

Fourier domain using convolution or filter back projection (FBP) methods [28].

The second class of reconstruction methods is algebraic. The basic principles of the algebraic reconstruction techniques (ART) are to deal directly with the set of linear equations similar to Eq. (1), in order to find the best values for the μ_{ij} coefficients. There are, however, two main difficulties. The first relates to the fact that the set of equations may be either underdetermined or overdetermined when the number of measurements does not match exactly the number of unknown values. The second relates to the large size of the matrix to be inverted.

These difficulties are partly minimized by the method developed by Kaczmarz [30] which is based on an iterative procedure. Initial values of μ_{ij} are chosen (usually taken as uniform inside the specimen and zero outside it). An iterative procedure is then applied in which the next estimate of the μ_{ij} values are calculated from the difference between the deduced and the measured intensities i_{mk}. The sparse structure of the μ_{ij} matrix added to the addition of constraints (such as the need for μ_{ij} to be positive and below a maximum value) allows us to reduce the number of operations at each iteration. After a few (3–5) complete passes through the image, the iterations are stopped when the μ_{ij} uncertainties fall inside those resulting from the statistical fluctuations of the signals [31, 32]. The artefacts of ART mainly concern the sharp edges of the object, like corners that are somewhat distorted because they are sampled by fewer equations; uniform areas may also show small variations.

The above descriptions of the image reconstruction methods was implicitly limited to the case of the reconstruction of a cross-section of the object illuminated by parallel beams. In industrial and medical CT with linear detectors, the projection data are mainly collected using fan-beams so that it is necessary either to derive the divergent beam reconstruction

algorithms (back projection methods) directly but modifying the inverse Radon transform formulas or, first, to deduce the expected projection given by a parallel illumination from the experimental projection acquired using the fan geometry (rebinning process).

If, as in X-ray microtomography using synchrotron radiation, a three-dimensional object is scanned by a parallel beam, it can be reconstructed from a set of two-dimensional slices each of which can be reconstructed using the foregoing algorithms. Such an operation is easy when the rotation axis is perpendicular to the X-ray beams but it is necessary to generalize the Radon transform theory when this is not the case [28].

When 2D detectors are used in combination with a point source, one has to deal with projections resulting from a cone beam geometry and specific algorithms have to be built to take into account the oblique ray paths through the specimen except for the midplane perpendicular to the rotation axis and defined by the X-ray source and one row of detectors where the fan beam methods can be applied.

3.3.2 Practical Solutions

The above methods only concern the image reconstruction from projections acquired with a perfect instrument. Unfortunately, there are some mechanical parameters influencing the resolution of the reconstructed images, such as the geometrical errors induced by the object rotation. These practical errors are due to various factors such as the misalignment of the specimen holder axis with respect to its ideal orientation (this axis describing cones during its rotation) and also the play of the movements transmitted by the stepped motor [33]. In microtomography these errors have first to be mechanically minimized and the remaining errors can be corrected by the reconstruction algorithm so that these errors fall below the resolution (a few micrometers) of the initial projections.

When classical X-ray sources are used there is also the beam hardening effect associated with the use of non-monochromatic radiation (see Sec. 2.1 of this Chapter). For quantitative X-ray (micro)-tomography, this spurious effect can be first reduced in the data acquisition by using, for instance, balanced filters [34] but it can also be minimized by taking into account the incident spectral distribution in the data processing. In the future, this effect will be canceled by the use of analytical CCD scanners

From the above analysis we find that a variety of instrumental architectures can be combined with a similar variety of the image reconstruction methods, even if the main methods used are mainly from the back projection class.

There are a number of algorithms for 3D reconstruction of the object from 2D projection images. For the cone beam geometry it is the Feldkamp's algorithm based on convolution and backprojection [35], an algorithm improved by using a triple cone beam reconstruction system by Lin et al. [36] and the Grangeat's algorithm based on the computation and the inversion of the first derivative of the Radon transform [37]. For the same geometry but using an algebraic reconstruction technique there are the algorithms proposed by Ham et al. [34] and by

Zolfaghari and Trebbia [38]. For parallel beam geometries of synchrotron X-ray microtomographs, the tomographic reconstructions are based either on the use of a direct Fourier inversion (DFI) algorithm [9] or on a more widespread Fourier filtered back projection algorithm [20].

Essentially, all the major algorithms are now available in the Reconstruction Lawrence Berkeley Laboratory Library [39], but the choice of a given algorithm is governed by the following points: (i) spatial resolution, (ii) artefacts, (iii) noise and contrast sensitivity, and (iv) the computation time for the image reconstruction.

The need for filtering, required when inverting the Radon transform and also for minimizing the artefacts associated with the finite number of projections, emphasizes the high spatial frequencies. But as practical projections are noisy, the choice of the reconstruction filter for minimizing the artefacts cannot be combined with filters for minimizing the amplification of noise (the parameter B in Eq. (3)) [28]. Also, the use of filters necessarily modifies the information contained at some spatial frequencies and the corresponding degradation of the spatial resolution. These remarks explain the variety of approaches mentioned above for finding the best compromise between the performance associated with points (i), (ii), and (iii), a compromise which depends upon the exact nature of the specimen and the important information to be extracted from its reconstruction [9].

In addition, there is point (iv). Conventional single slice medical CT scanners require extensive computational resources. Using an array processor (running at $\approx 10^7$ floating-point operations per second) and a standard reconstruction algorithm (FBP) more than one month would be needed to invert projection data for a 3D sample sectioned into 1000 projections each composed of $10^3 \times 10^3$ pixels. Thus, without faster algorithms or higher speed computers, tomography in this format would be impractical [9]. Methodologies based on direct Fourier inversion (DFI) offer advantages in speed mainly with respect to algebraic reconstruction techniques which are more time-consuming for the inversion of a rather large matrix.

Associated with the reconstruction processes is the problem of displaying a 3D object on a 2D screen. From the stack of 2D section images, software routines are available to assemble these slices into a block of data that can be manipulated as a three-dimensional object. This object can then be viewed from different angles and cross sections in arbitrary planes, while graphics workstations also support more complex operations [6, 40].

Starting from projections acquired with the same laboratory microtomograph [41], Fig. 6 illustrates two different approaches for the data reconstruction algorithm and display. Figure 6a shows one of the 67 projected images of a diode, each image being acquired in 5 s by the use of a CCD camera composed of 256×256 pixels. From each of these projections a specific 1D projection line has been selected to obtain the corresponding cross section of the object by a backprojection process of this slice and the process has been repeated for other slices. Figure 6b represents one of the reconstructed cross sections (across a square silicon chip) which has been performed in less than 30 s using a back projection algorithm and a 486/50 MHz computer [33]. Figure 6c represents the reconstructed image of a plastic (carbon

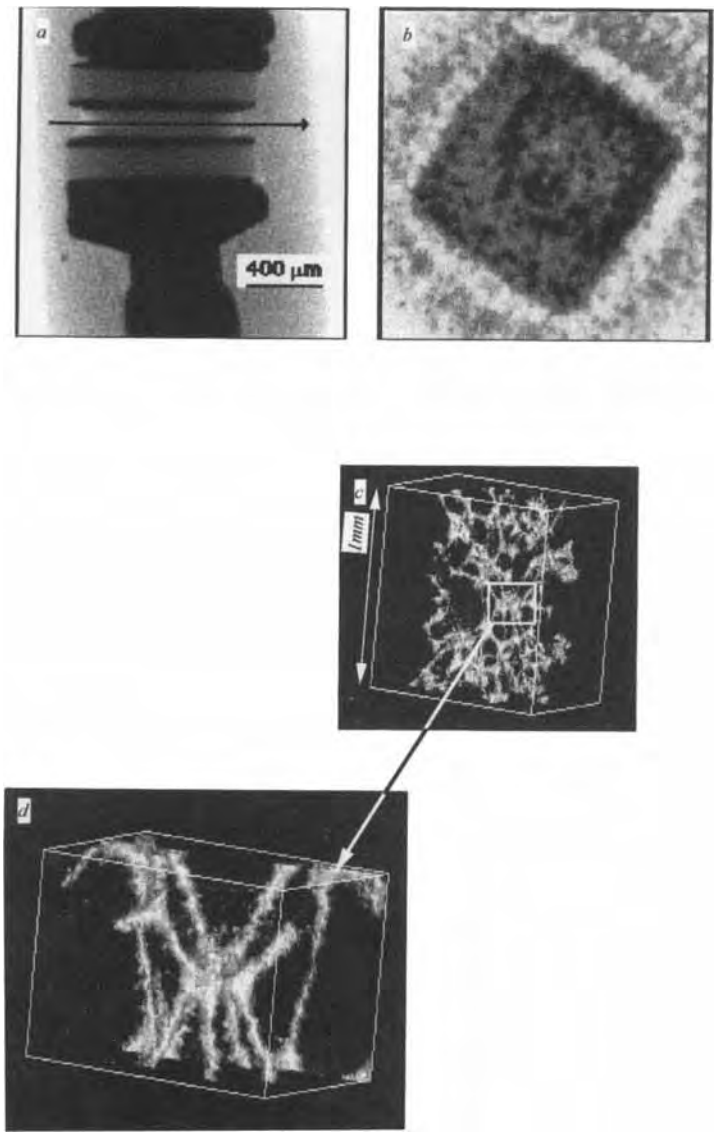


Figure 6. Examples of 3D reconstruction from projections acquired by using laboratory equipment [18, 41]. (a) X-ray projection image of a diode and (b) reconstructed cross-section showing the square shape of the silicon block (back projection method) [33]. (c) 3D reconstructions of a plastic foam (ART method) [38]. (d) is the enlarged view of (c).

based) foam also obtained from 67 projections (1 min each) but using an ART algorithm (five complete iterations). The typical dimension of the object is one millimeter divided into 125^3 voxels and the computation time was around 4 h with a UNIX workstation SUN SPARC 2 for the complete reconstruction of the object [38].

3.4 Conclusions

Microtomography can be used to non-destructively generate images of internal sections of an opaque object of 1–10 mm dimensions with resolution in the 3–30 µm range; no special sample preparation is required. Associated with the development

of high speed and efficient methods for the image reconstruction from projections, the dramatic progress of this technique has resulted from the use of modern two-dimensional detectors for the rapid acquisition of the initial digital images. This technique can operate using synchrotron sources of radiation or shadow X-ray microscopes derived from elementary modifications of an SEM or conventional microfocus X-ray sources. Recent technical developments are concerned with quantitative X-microtomography using monochromatic projections and elemental mapping in three dimensions. Presently these developments are mainly restricted to microtomographs operated with S.R. sources [42–43] but the progress in analytical CCD cameras may open these possibilities to microtomographs operated with conventional X-ray sources [44–47]. Another technical development consists in improving the lateral resolution. Like in electron microscopy (see Sec. 8.2 of Chap. 8, this Volume) this improvement can be obtained by focusing the X-ray beam on flat and thin specimens recording next a small number of two-dimensional projections over a reduced angular range. From such a partial reconstruction, a submicrometer resolution has been recently reached using soft X-rays on a test specimen, five microns thick [48]. From the point of view of applications, X-ray tomographic microscopy is beginning to be applied to the investigation of composite materials [49–51]. They are also concerned with dynamical studies such as the diffusion of liquids into porous materials [51]. Due to the sensitivity of X-rays for the detection of medium weight elements embedded in light matrices (see Sec. 2.1 of this Chapter) the fields of applications of this emerging technique cover not only engineering materials and catalysts but also geology [52] (see also Vol. II, Chap. 1, Sec. 2) and life sciences [53].

3.5 References

[1] G. N. Hounsfield, *Br. J. Radiol.* **1973,** *46*, 1016 (with A. M. Comack).

[2] P. New, W. Scott, J. Schmur, K. Davis, J. Taveras, *Neurobiology* **1974,** *110*, 109.

[3] P. K. Hunt, P. Engler, W. D. Friedman, *Advances in X-ray Analysis* **1988,** *31*, 99 and most of the papers in the same issue such as that of J. C. Russ, p. 25, Ryon et al., p. 35.

[4] G. T. Herman, *Image Reconstruction from Projections*, Academic Press, New York, **1980.**

[5] J. C. Russ, *Computer-assisted Microscopy*, Plenum Press, New York, **1990,** Chap. 13, p. 419.

[6] D. C. Copley, J. W. Eberhard, G. A. Mohr, *J. Miner. Met. & Mat.* **1994,** *46*, 14.

[7] D. A. Chesler, R. S. Riederer, J. J. Pelc, *J. Comp. Ass. Tomogr.* **1977,** *1*, 64.

[8] L. Grodzins, *Nucl. Instrum. & Methods* **1983,** *206*, 541.

[9] B. P. Flannery, H. W. Deckman, W. G. Roberge, K. L. D'Amico, *Science* **1987,** *237*, 1439; B. P. Flannery, W. G. Roberge, *J. Appl. Phys.* **1987,** *62*, 4668.

[10] A. Y. Sasov, *J. Microsc.* **1987,** *147*, 169 (part 1) and 179 (part 2).

[11] J. Cazaux, J. Despujols, Traité d'Analyse Chimique et Caractérisation, Les techniques de l'Ingénieur, Paris, **1992,** *7*, Part 950, 1.

[12] N. Gurker, R. Nell, W. Backfrieder, J. Kandutsch, K. Sarg, S. Prevrahl, C. Neutwich, *Nuclear Instr. & Methods Phys. Res.* **1994,** *B94*, 154.

[13] J. C. Elliott, P. Anderson, X. J. Gao, F. S. L. Wong, G. R. Davis, S. E. P. Dowker, *J. X-Ray Sci. & Technol.* **1994,** *4*, 102.

[14] P. Wells, J. R. Davis, B. Swendermann, P. A. Stadbolt, N. Bena, J. A. Grant, D. R. Davis, M. J. Morgan, *Nuclear Instr. & Methods Phys. Res.* **1992,** *B72*, 261.

[15] S. Aoki, I. Yoneda, T. Nagai, N. Ueno, K. Murakami, *Jpn. J. Appl. Phys.* **1994,** *33*, L556.

[16] M. Marsl, H. Reiter, P. Möller, *J. Engineer. Mat. Techn.* **1990,** *112*, 223.

[17] P. C. Cheng, D. M. Shinozaki, T. H. Lin, S. P. Newburry, R. Sridhar, W. Tarng, M. T. Chen, L. H. Chen in *X-ray Microscopy III, Springer Series in Optical Sciences* **1992,** *67*, 184.

[18] J. Cazaux, D. Erre, D. Mouze, J. M. Patat, S. Rondot, A. Sasov, P. Trebbia, A. Zolfaghari, *J. Phys. IV (France)* **1993,** *C7*, 2099.
[19] T. Hirano, S. Eguchi, K. Usami, *J. Appl. Phys.* **1989,** *28*, 135.
[20] J. H. Kinney, D. L. Haupt, M. C. Nichols, T. M. Brunig, G. N. Marshall, Jr, S. J. Marshall, *Nucl. Instr. & Meth. Phys. Res. A* **1994,** *347*, 480.
[21] J. H. Kinney, Q. C. Johnson, R. A. Saroyan, M. C. Nichols, U. Bonse, *Rev. Sci. Instr.* **1989,** *60*, 2475.
[22] U. Bonse, R. Nubhardt, F. Busch, J. M. Kinney, R. A. Saroyan, M. C. Nichols in *X-ray Microscopy III, Springer Series in Optical Sciences* **1992,** *67*, 167.
[23] H. W. Deckman, J. H. Dunsmuir, K. L. D'Amico, S. R. Ferguson, B. P. Flannery, *Mat. Res. Soc. Symp. Proc.* **1991,** *217*, 97.
[24] U. Bonse, F. Busch, O. Gunnewig, F. Veckmann, R. Pahl, G. Delling, M. Hahn, W. Graeff, *Bone and Mineral* **1994,** *25*, 25.
[25] J. C. Elliott, S. D. Dover, *J. Microsc.* **1982,** *126*, 211.
[26] R. Nusshardt, U. Bonse, F. Busch, J. H. Kinney, R. A. Saroyan, M. C. Nichols, *Synchr. Rad. News* **1991,** *4*, 21.
[27] J. Radon, *Math. Phys. Klass Leipzig* **1917,** *69*, 262.
[28] A. K. Jain in *Fundamentals of Digital Image Processing*, Prentice-Hall, Englewood Cliffs, NJ, **1989,** Chap. 10, p. 431.
[29] R. M. Lewitt, *Proc. IEEE* **1983,** *71*, 390 and other papers of this special issue (no. 3) on computerized tomography (pp. 293–431).
[30] S. Kaczmarz, *Bull. Acad. Pol. Sci. Lett. A* **1937,** *35*, 355.
[31] R. Gordon, *IEEE Trans. Nucl. Sci.* **1974,** *NS 21*, 70.
[32] Y. Censor, *Proc. IEEE* **1983,** *71*, 409.
[33] A. Y. Sasov, J. Cazaux in *X-ray Microscopy IV* (Ed: V. V. Aristov and A. I. Erko) Bogorodski Pechatnik Pub. Co., Chernogolovka, Russia **1994,** p. 157.
[34] Y. S. Ham, J. C. Russ, R. P. Gardner, K. Verghese, *Appl. Rad. Isot.* **1993,** *44*, 1313.
[35] L. A. Feldkamp, L. C. Davis, J. M. Kress, *J. Opt. Soc. Am.* **1984,** *1A*, 612.
[36] T. H. Lin, G. Wang, P. C. Cheng in *X-ray Microscopy III, Springer Series in Optical Sciences*, Springer-Verlag, Berlin, **1992,** *67*, 296; G. Wang, T. H. Lin, D. M. Shinozaki, H. G. Kim, P. C. Chen in *Multidimensional Microscopy* (Ed: P. C. Chang, T. H. Lin, W. L. Hu, J. L. Wu), Springer-Verlag, New York, **1993,** Chap. 9; G. Wang, T. H. Lin, P. C. Cheng, D. M. Shinozaki, *IEEE Trans. Medical Imaging* **1993,** *12*, 486.
[37] P. Grangeat, P. Sire, P. Le Massus, P. Mellenec, P. Rizo, *SPIE, Medical Imaging V*, Image Verlag, Berlin, **1991,** p. 320.
[38] A. Zolfaghari, P. Trebbia in *X-ray Microscopy IV* (Ed: V. V. Aristov and A. I. Erko) Bogorodski Pechatnik Pub. Co., Chernogolovka, Russia **1994,** p. 439. A. Zolfaghari, Thèse de Doctorat Université de Reims, **1994.**
[39] Huesman, *User's Manual: Donner Algorithms for Reconstruction Tomography*, PUB-214, Research Medicine Group, Donner Laboratory, Lawrence Berkeley Laboratory, University of California, Berkeley, CA, **1994,** p. 720.
[40] J. C. Russ, *The Image Processing Handbook*, 2nd ed., CRC Press, Boca Raton, FL USA, 547 and 589, **1995.**
[41] J. Cazaux, H. El Hila, D. Erre, D. Mouze, J. M. Patat, S. Rondot, P. Trebbia, A. Zolfaghari, *European Microscopy and Analysis* **1994,** May, p. 19; Vol. 29.
[42] X. Y. Wu, F. A. Dilmanian, Z. Chen, B. Ren, D. N. Slatkin, D. Chapman, M. Shleifer, F. A. Staicu, W. Thomlinson, *Rev. Sci. Instrum.* **1995,** *66*, 1346.
[43] P. Spanne, K. W. Jones, *Rev. Sci. Instrum.* **1995,** *66*, 2377.
[44] R. A. Stern, K. Liewer, J. R. Janesick, *Rev. Sci. Instrum.* **1983,** *54*, 198.
[45] J. P. Doty, G. A. Luppino, G. R. Ricker, *S.P.I.E.* **1986,** *688*, 216.
[46] H. Tsunemi, K. Mizukata, M. Hiramatsu, *Jap. J. Appl. Phys.* **1988,** *27*, 670.
[47] J. Janesick, A. Collins, D. Blake in *Microbeam Analysis* (Ed.: E. S. Etz), Weinheim VCH **1995,** p. 7.
[48] I. McNulty, W. S. haddad, J. E. Trebbes, E. H. Anderson, *Rev. Sci. Instrum.* **1995,** *66*, 1431.
[49] T. M. Breunig, J. C. Elliott, S. R. Stock, P. Anderson, G. R. Davis, A. Guvemlir in *X-ray Microscopy II, Springer Series in Optical Sciences*, Springer-Verlag, Berlin, **1992,** *67*, 465. P. M. Mummery, P. Anderson, G. R. Davis, B. Derby, J. C. Elliott, *Scr. Metall. Mater.* **1993,** *29*, 1457.
[50] G. Y. Backlini, R. T. Bhatt, A. J. Eckel, P. Engler, M. G. Castelli, R. W. Rauser, *Materials Evaluation* **1995** (Sept.) p. 1040.
[51] H. Kinney, T. M. Breunig, T. L. Starr, D. Haupt, M. C. Nichols, S. R. Stock, M. D. Butts, R. A. Saroyan, *Science* **1993,** *260*, 789.
[52] J. K. Jasti, G. Jesion, L. Feldkamp, SPE: Society of Petroleum Engineers, *Formation Evaluation*, **1993,** Sept. p. 189, J. K. Jasti, G. Jesion, L. Feldkamp, SPE: Society of Petroleum Engineers, **1990,** no. 20495, p. 269.
[53] X. J. Gao, J. C. Elliott, P. Anderson, G. R. Davis, *J. Chem. Soc., Faraday Trans.* **1993,** *89*, 2907.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

4 Soft X-Ray Microscopy by Holography

4.1 Introduction

Soft X-ray microscopy can be performed by various methods that are described elsewhere in this book. In the present state of the art, the best resolution (50–30 nm) is obtained with systems based on diffractive lenses (Fresnel zone plates, FZPs), working in transmission, whereas a better collecting efficiency of the flux might be obtained in the near future with high-reflection mirror optics, working under normal incidence, owing to the progress in multilayer coatings.

Despite these successes, Gabor holography is still a possible alternative to other methods, because it seems to offer three decisive advantages:

- three-dimensional (3D) imaging capability, that is, in-depth resolution r_L close to the transverse resolution r_T;
- no X-ray imaging component needed in the recording step;
- simultaneous recording of the absorption and phase of the object transmission.

It is necessary to point out at the very beginning of this article that the first point is not effective in the present context of X-ray microscopy. This is related to the fact that, with holographic imaging as with any other imaging method, the *imaging aperture* governs both the transverse resolution r_T and the in-depth one r_L, yielding between these two parameters the relation $r_L = 2r_T^2/\lambda$, where λ is the X-ray wavelength. Even assuming a transverse resolution $r_T = 0.05\,\mu m$ (the best resolution ever obtained, see Sec. 4.4 of this Chap.), one finds $r_L = 1.4\,\mu m$ at $\lambda = 3.5\,nm$, which is 30 times greater than the transverse resolution. This is therefore far from being a real and useful 3D capability, and the situation is even worse at shorter wavelengths.

The next claimed advantage, the lack of an X-ray component, is still a true advantage, because high-resolution X-ray optical components are quite difficult to realize, and, at least for FZPs, further progress appears to be more and more difficult. Moreover, the microscopes are quite complicated devices, where fine adjustments, stability requirements, etc., require a lot of time in order to get the desired image.

Finally, recording both the phase and amplitude of the sample is still attractive, as the recent trend in microscopy of biological samples is to minimize the radiation dose deposited into the sample, in order to preserve their integrity. Phase imaging is therefore the method of choice because it allows working in spectral regions where absorption is small.

The history of Gabor holography as a possible X-ray microscopy method is almost as old as Gabor holography itself. As early as 1952, Baez [1] made the transposition from Gabor's original idea (which applied to electron microscopy). At that time, no X-ray diffracting components existed, and grazing incidence mirror optics had a quite limited resolution. Further theoretical investigations followed [2, 3], until synchrotron radiation (SR) sources became available in the 1980s, which allowed the whole field of practical X-ray microscopy to develop quickly. After early work by Giles [4] and Aoki [5], the recording of Gabor X-ray holograms for the purpose of microscopy was first demonstrated by Howells et al. [6, 7], followed by Joyeux et al. [8]. Later, the use of Fourier transform holography for microscopy was demonstrated by McNulty et al. [9].

We now examine in some detail the basics and the practical realization of X-ray microscopy by Gabor holography. Pseudo-Fourier holography will be discussed later. As the purpose of this article is not to give the theory of holography in general, we shall concentrate on those points that are specific or critical to X-ray microscopy.

4.2 Basics of Gabor Holography

Holography is a two-step process: first *record* then *reconstruct* the hologram (Fig. 1). In the *recording step*, an interferogram is created and recorded (usually on a photochemical medium) that carries all the pertinent information concerning the phase and amplitude repartition of the object wavefront. This always implies using a second wavefront, the reference wavefront, which is added coherently to the object wavefront to realize the interferogram. In the *reconstruction step*, a replica of the original wavefront is created by allowing the hologram to diffract a suitable 'reconstruction wavefront' (which can be different in shape and wavelength from the reference wavefront). According to Huyghens's principle, this allows us to 'see' through the replica the objects that gave rise to the original recorded wavefront.

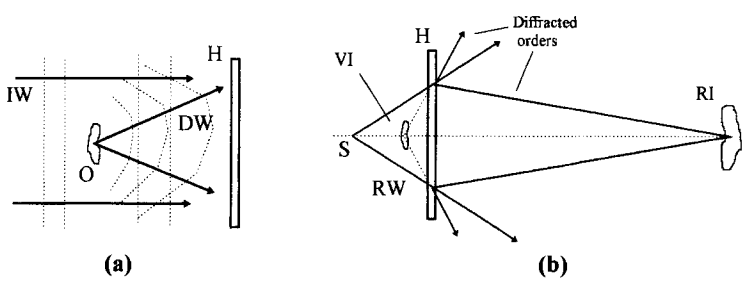


Figure 1. The principle of Gabor holographic microscopy. (a) Recording step. (b) Optical reconstruction step, yielding one reduced virtual image and one magnified real image. The magnification does not arise solely from the wavelength change: the reconstruction wavefront shape must be different from that of the reference wavefront (e.g., spherical versus plane). O, object; H, hologram; IW, DW, RW, incident, diffracted, and reconstruction wavefronts and rays; S, reconstruction source; VI, RI, virtual and real images.

In the case of so-called in-line or Gabor holography, a weakly absorbing object is illuminated by a coherent wavefront, and the Fresnel diffraction pattern produced downstream along the transmitted beam is the Gabor or in-line hologram of the object. In fact, the emergent light can be said to consist of two parts: (1) an unperturbated transmitted wavefront (undeviated rays); and (2) a diffracted part, due to the variations of the object transmittance. If the incident wavefront is coherent enough (see below), both parts add coherently, that is, they interfere. According to the general definition, the resulting pattern is the hologram of the modulated part of the object, with the unmodulated part acting as a reference. This implies that the object plane must be weakly modulated, or at least sparsely filled, in order that nowhere in the hologram plane does the reference vanish.

4.3 Recording a Gabor X-Ray Hologram

Various points must be addressed before setting up the recording experiment: (1) the coherence conditions of the incident wavefront; (2) the average hologram illuminance, in order to get a reasonable recording time; (3) the resolving power of the recording medium; and (4) the practical realization of the set-up, which must meet some stability requirements. All these points are mainly governed by the final resolution that is desired at the end of the imaging process.

To perform the analysis, we consider a single ideal object point, that is, we proceed in terms of the point spread function. The different relevant parameters are outlined in Fig. 2. The source is placed on the axis at a distance D, much greater than the distance d from the object to the recording plane, so that the incident wavefront can be said to be plane in the transverse extension of the set-up. Dimensions in a typical implementation are D of several meters and $d < 2\,\text{mm}$. α is the deviation angle of a diffracted ray, which we always assume to be small.

In order to get a final resolution r in the reconstructed image, obviously the corresponding information must be first recorded: we may speak thus of this as the recorded resolution. Therefore, using a Shannon criterion, this means that interference must take place and be recorded up to the diffraction angle $\alpha = \lambda/2r$, where λ is the X-ray wavelength. Here, α is simply a collecting aperture, as in any optical instrument. Note also that the largest spatial frequency to record is therefore just the same as the diffracting one.

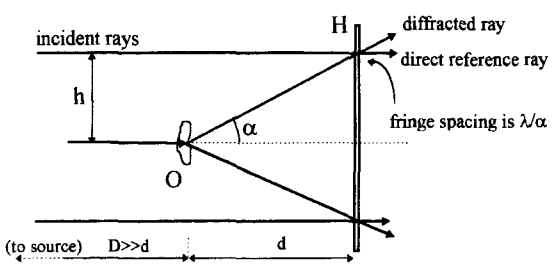


Figure 2. The relevant geometrical parameters of the recording step.

4.3.1 Coherence Conditions

For that limiting ray, the separation between the reference and object rays is $h = \alpha d = \lambda d/2r$, and the path difference at the interference point is $\Delta = \alpha^2 d/2 = \lambda^2 d/8r^2$, or, expressed in wavelengths: $N = \Delta/\lambda = \lambda d/8r^2$. Δ or N set the temporal coherence conditions, that is, the acceptable spectral width of the source: one ray split in two and delayed by N wavelengths must still interfere with its undelayed part. The condition to meet is that the relative spectral width of the source $\lambda/\Delta\lambda$ be notably smaller than $N = \lambda d/8r^2$. h sets the spatial coherence needed in the incoming wavefront: two rays separated by h must produce well-modulated interferences. As a numerical example, with $\lambda = 5\,\text{nm}$, $d = 2\,\text{mm}$, and $r = 0.1\,\mu\text{m}$, the coherence radius is $h = 50\,\mu\text{m}$ and $N = 125$.

It is important to note that h and Δ are relative to the object point to record, regardless of the whole object size: the coherence requirements are local. As the mutual coherence of a beam is also a local property, this means that Gabor holography allows the recording of a field which is as large as the incoming beam itself (even millimeters), regardless of the desired resolution and required coherence. This unique property of Gabor holography is one of its major practical interests, as it even allows several samples to be recorded at the same time, without the need to search and focus on a particular (and generally small) 'region of interest'. This task is postponed to the reconstruction step, outside of the X-ray facility context.

Although h and Δ set the coherence conditions to meet, one parameter remains free up to this point, namely the distance d. This parameter influences both steps of the imaging process. Firstly, both coherence constraints are increased for increasing d, which may also decrease the record illuminance when an incoherent source is used (due to the setting of the spatial coherence, see below). Secondly, it is well known that a Gabor hologram produces at least two reconstructed images ('twin images') at different places of the optical axis, as illustrated in Fig. 1, and the in-focus one is more or less blurred by the out-of-focus one. This blur can be decreased by increasing the image separation, which is precisely proportional to d, or, better, to N (here N plays the same role as the number of lines of a grating with respect to its spectral resolution). A reasonable choice is $N \approx 100$, but smaller values, down to $N = 20$ or 30, may be considered when a trade-off is necessary. Again using $r = 0.1\,\mu\text{m}$ and $\lambda = 5\,\text{nm}$, one finds $d = 1.6\,\text{mm}$ for $N = 100$.

4.3.2 Spatial Coherence and the Hologram Illuminance

If true X-ray lasers existed, that is, having a full transverse spatial coherence, as for the helium–neon laser, this section would have no use, and the record illuminance (i.e., the recording time) would be simply set by the beam power and its transverse size in the hologram plane. However, synchrotron radiation sources are still the best X-ray sources when some coherence is needed, and some of these sources are far from being full spatially coherent. In any case, the source coherence properties must be checked, and, if needed, the spatial and temporal coherence must be synthesized.

As already mentioned, the latter is easily obtained by spectral filtering, that is, using a monochromator to restrict the spectral width to the value allowed.

The question of spatial coherence and hologram illuminance is more complex, and it may be helpful to introduce two fundamental parameters of beam optics, namely the beam *étendue* and the coherent *étendue*. The étendue of a beam is the product of its normal section by its aperture (solid angle), seen from this section. It is an invariant of the beam during the propagation through an optical system (but with neither diaphragmation nor diffusion). It can be shown that due to diffraction, no beam can have an étendue smaller than $(\lambda/\beta)^2$, where β is a numerical coefficient. Moreover, for any pencil out of a propagating beam and having this 'coherent étendue', the rays inside the pencil are mutually coherent. The flux carried by a given étendue depends only on the spectral brilliance of the source (flux per unit of beam étendue per spectral unit). Therefore, the flux which is available in one coherent region of a beam is determined only by the beam spectral brilliance and the wavelength squared (i.e., the coherent étendue). Note that the illuminance also depends on the object to hologram distance d, as already mentioned.

A consequence of this classical result of photometry and coherence is that only SR sources are practical for the realization of high-resolution soft X-ray holography or other coherent experiments, because of their highly collimated beam, even if it is not spatially coherent. As an example, a typical beam divergence for a 'low-coherence' SR beam may be 1 mrad, with a typical linear transverse dimension of 100 μm; this is by far not a fully coherent beam at $\lambda = 5$ nm. Compared to the 2π st of a classical bremsstrahlung source, the gain in luminance (assuming the source surface and the flux radiated are the same) is as large as 5×10^5.

However, two other kinds of soft X-ray sources may become usable in the short or medium term, namely X-ray lasers and high harmonic-order generation. In particular, lasers may exhibit an increase of several orders of magnitude in luminance with respect to the best SR sources, even considering the small repetition rate of these sources.

4.3.3 Realization of the Coherence Conditions

From the previous Section, to optimize the recording set-up spatial coherence simply means the adjustment of the section of one coherent étendue, in order to fill the required size—and no more—in the hologram plane.

In practice, as long as a monochromator is needed to adjust the spectral width, the best solution is to use the exit slit (considered as an incoherent source) as the holographic source. The section of its spatially coherent étendue in the hologram region may be adjusted by changing the slit size, with the help of the classical Zernike–Van Cittert theorem [10]. But this also changes the spectral width. Obviously, an optimization must be carried out on the monochromator dispersion, in order to keep the different coherence terms to just what is needed.

As a detailed analysis is outside the scope of this article, only the main result will be given. Because the global coherence

is the product of the temporal and spatial term, the coherence has to be split between these terms. Thus, given the total coherence and the maximum object resolution r, the system optimization leads to a fixed monochromator dispersion (K, exit slit width/spectrum width): $1/K = 2r/D$. With $r = 0.1\,\mu m$ and $D = 10\,m$, $K = 50\,mm\,nm^{-1}$, that is, 0.02 nm spectrum per millimeter slit. Remember that although the dispersion optimum value does not depend on the distance d the optimal hologram illuminance value does, through the size of the coherence zone.

4.3.4 The Recording Medium

As stated earlier, the recording medium must take account of the following requirements: sensitivity to X-rays, very high resolution, and the ability to read the hologram in the reconstruction step. The main constraint is indeed the resolution constraint, since recording for a final 0.1 μm implies recording to more than $5000\,mm^{-1}$. Most workers in the field use microlithographic resists (with sensitivity to either electrons or energetic photons), which are available off the shelf. These have the additional advantage of producing a relief hologram, which is much easier to read out in the reconstruction step.

4.3.5 About the Practical Implementation

The main constraint of the practical realization (holographic camera) is the mechanical stability of the holographic plate, whatever it is, with respect to the fringes. It is enhanced by the fact that the whole set-up from the slit source to the hologram is usually in vacuum, as soft X-rays are very quickly absorbed in air (helium could be used for millimetre distances). That stability is scaled by the fringe spacing. In principle, instability may occur due to thermal drift or vibrations. It is often stated that this is a drawback of holographic microscopy. This is not true, however, because the same kind of mechanical stability is required in any imaging instrument having the same resolution. Moreover, the extreme simplicity of the Gabor holographic set-up (Fig. 3) makes it much more simple to get a stable assembly than in an imaging device, in which moving parts are needed to perform the focusing adjustments.

The stability of the source with respect to the holographic set-up is also necessary, because any transverse displacement of the source is equivalent to an increase in its width, that is, to decrease the spatial coherence on the hologram. Therefore, the scale to consider is the source width itself. This is an argument to the choice of a large value for D, the source-to-hologram

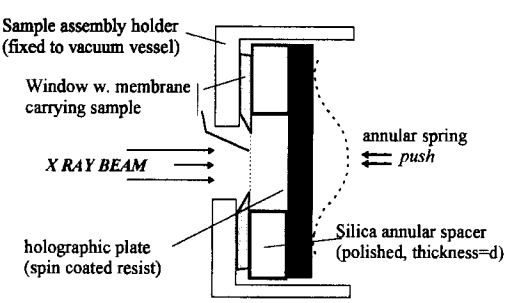


Figure 3. A possible realization of the recording setup. All the system is in vacuum. Typical distances are: spacer thickness: $d = 0.5$–2 mm; field (i.e. membrane extension): 1–4 mm; membrane thickness: 100–500 nm.

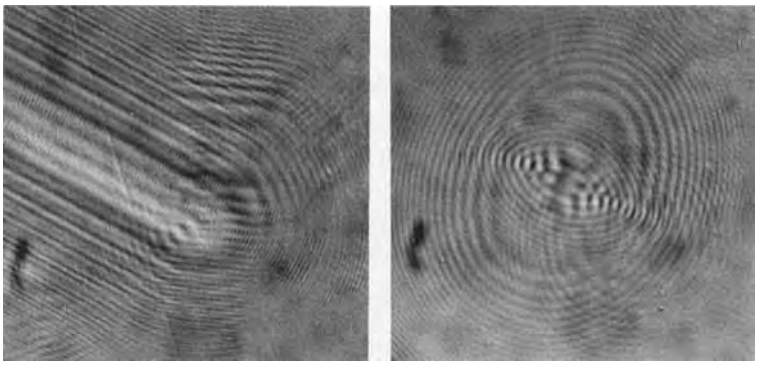


Figure 4. Two typical Gabor holograms of diatoms (the classical 'test' of microscopists). They are relief patterns recorded at $\lambda = 10\,\text{nm}$ (with $d \approx 5\,\text{mm}$) in a lithographic photoresist. These images are interferential contrast micrographs. The field is $100 \times 100\,\mu\text{m}$.

distance. In a typical case, using $D = 10\,\text{m}$, $r = 0.1\,\mu\text{m}$, and $d = 2\,\text{mm}$ implies a slit width greater than $200\,\mu\text{m}$, which makes the stability easy to realize.

Finally, an important component of X-ray microscopy is the sample-holding membrane. As soft X-ray absorption in matter is total within a few micrometers or less, it is necessary to hold samples on very thin membranes, with reasonable transverse dimensions. Such membranes, which were quite tricky to realize in the past, are now available on order. Most classical ones are Si_3N_4 on silicon windows. Thicknesses as small as 100 nm with a millimetric aperture can be obtained. Other possible materials are mainly silicon SiO_x, and also some polymers.

4.3.6 Further Comments

Despite the common belief that the recording step (stricto sensu) of an X-ray hologram is not the most difficult part of the holographic imaging process, once a suitable X-ray facility and photon-line arrangement are available (and this indeed requires a lot of technology). Probably the most critical part of the experiment is the preparation and development of the resist-coated holographic plates, because high cleanliness standards have to be met. Figure 4 shows micrographs of such an X-ray hologram.

In fact, in the author's opinion, part of the advantage of holographic microscopy in the soft X-ray domain is related to the fact that very little technology, and in particular no X-ray focusing technology, is involved in the Gabor holographic process (excluding, of course, the source technology). This makes it completely free of defects due to X-ray components, and progress toward higher recorded resolution depends only on the recording medium (a lot of work is being done on high-resolution lithographic resists), and the source luminance, on which recent progress has been impressive. However, the reconstruction step must also accommodate this high resolution.

4.4 Reconstruction of Gabor Holograms

Once recorded, a hologram must be 'reconstructed', in order to decipher the

recorded information into an ordinary observable image. This step is the difficult one, in the case of X-ray microscopy.

The reconstruction process consists, in principle, of applying a propagation operator to the recorded pattern, interpreted in terms of the complex amplitude of a wavefront. This can be done either by computer or optically. In both cases, a major difficulty is the so-called twin image, already mentioned in this article. Only computational solutions may be applied to decrease the twin image blur. This is an important topic in itself, which will not be developed further here (interested readers are directed to Koren et al. [11]).

4.4.1 Reconstruction by Computer

For the purpose of illustration, let us consider a $20 \times 20\,\mu m$ object field to image with $0.1\,\mu m$ resolution (i.e., 200×200 pixels). We assume a recording coherence size (i.e., the size of the hologram of one single point; see Sec. 4.2 of this Chapter) of only $40\,\mu m$. The hologram field is therefore $60\,\mu m$, and it must be read with a pixel size much smaller than $0.1\,\mu m$, say $0.04\,\mu m$ or better, that is, about 1500×1500 well-resolved pixels.

Speaking of the computation itself, such an image size is large but acceptable. The main problem of computer reconstruction is the necessary read-out of the photochemical record into a numerical record, because a great number of very small pixels have to be read with high dimensional accuracy. The position accuracy should probably be much better than the pixel size (at least within the extension of the hologram of a single point), in order to introduce negligible phase error in the computation.

This approach has been demonstrated by Jacobsen et al. To perform the hologram read-out they used a transmission electron microscope, after shadowing of the resist relief hologram. A resolution of about $0.1\,\mu m$ was obtained [12]. However, as the shadowing may produce phase distortion, this group proposed and tested [13] the use of an atomic force microscope, with enhanced positioning accuracy. They have recently demonstrated a resolution of $0.04\,\mu m$ with a biological sample (Fig. 5). It is believed that smaller resolutions can be reached, because there is no theoretical limit to the final resolution, except possibly the number of X-ray photons received by a pixel during the recording [14].

4.4.2 Optical Reconstruction

In the present context, the goal of optical reconstruction is to realize a specific instrument for hologram reconstruction which is as easy to use as any modern optical microscope, such as a confocal microscope.

4.4.2.1 Basics

However, this approach implies using an 'optical' wavelength, that is, to which ordinary optical techniques apply (i.e., $\lambda_c > 0.15\,\mu m$), much larger than the recording X-ray wavelength. This obviously sets a limit to the theoretical best resolution which can be obtained in the final image. Moreover, if that basic limit is to be reached, a large optical

aperture is necessary, and it is well known that huge optical aberrations will have to be eliminated. Finally, it is also well known that the optical noise is a severe drawback in the optical reconstruction of Gabor holograms. In summary three points must be addressed: (1) What is the theoretical limit of resolution? (2) How can this limit be reached and the aberrations eliminated? (3) How can most of the optical noise be got rid of?

Before answering these questions, let us remember the principle of optical reconstruction. To this end, we consider the hologram of two object points ('point holograms') separated by ρ (Fig. 6). Both point holograms are identical Fresnel zone patterns, just shifted from each other by ρ (provided the recording source was far enough from the sample). It is assumed that no resolution limitation occurs from the recording step, that is, the recording aperture is large enough, so that the recording step is not the limiting one. Reconstruction is usually made by illuminating the hologram by a spherical wave, from a source situated at a distance z in front of the hologram plane. We assume that the optical axis is defined by the source S and the center of one of the two

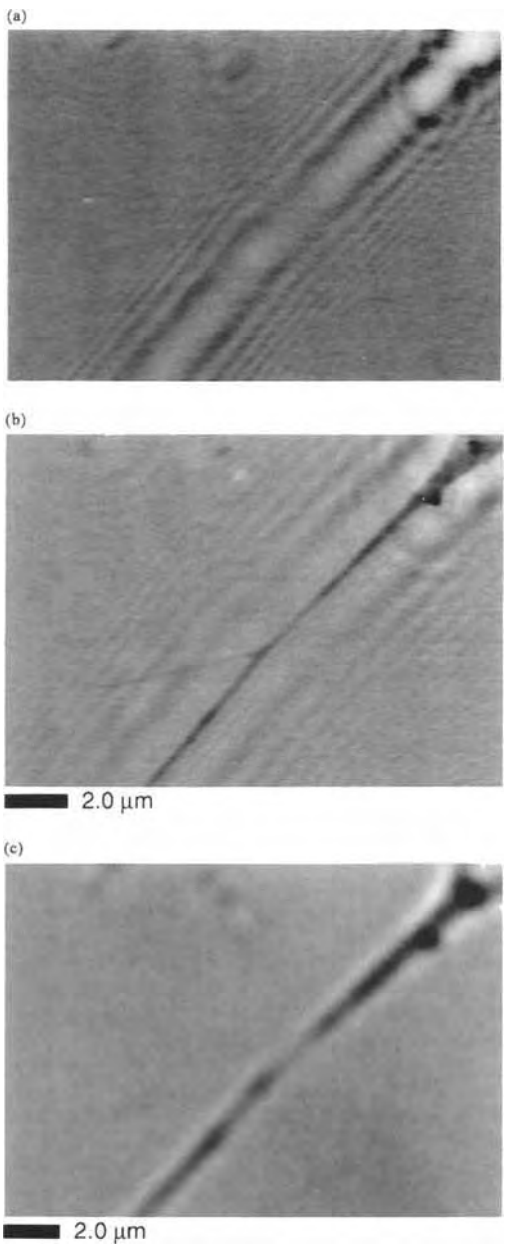


Figure 5. (a) X-ray Gabor (in-line) hologram and (b) numerically reconstructed phase contrast image of a tendril from an NIL8 hamster neural fibroblast. The hologram was recorded on photoresist using a coherent beam of soft X-rays, $\lambda = 1.89$ nm, from an undulator source. A linear-field atomic force microscope was used to read the topography of the developed photoresist, and the digitized hologram was then reconstructed numerically. (c) Following hologram exposure, the specimen was also imaged by optical microscopy using a 0.90 NA dry lens suitable for this critical-point dried sample. Intercellular organelles are seen more clearly in the reconstructed X-ray hologram, and the tendril at the lower left is clearly visible in the X-ray hologram while it can not be seen in the optical micrograph. Using an atomic force microscope for hologram examination and numerical reconstruction has yielded close to 40 nm resolution. (Courtesy of S. Lindaas and M. Howells (Lawrence Berkeley National Laboratory) and C. Jacobsen and A. Kalinovsky (SUNY at Stony Brook).)

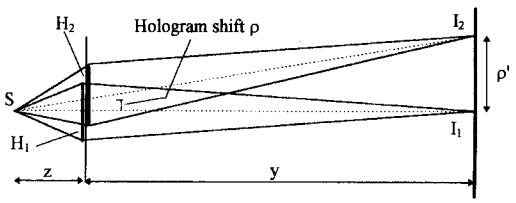


Figure 6. Principle of the optical reconstruction set-up. H_1 and H_2 are the holograms of two object points separated by ρ. The *optical magnification* is $M_O = y/z$; the *holographic magnification* is $M_H = \rho'/\rho = 1 + M_O$.

point holograms, so that one of them is on-axis. According to the conjugation laws in holography, each point hologram acts as a lens with focal length $f_c = \lambda d/\lambda_c$, which focuses the light toward one image point at a distance y. This also shows that the reconstruction parameters are determined by the product (λd) only, once λ_c is fixed. Orders of magnitude should be in mind: with $\lambda = 5\,\text{nm}$ (soft X-rays), $\lambda_c = 0.25\,\mu\text{m}$ (ultraviolet), and $d = 1.8\,\text{mm}$, the focal length f_c is $36\,\mu\text{m}$.

4.4.2.2 The Limit of Resolution

In ordinary optics, the image resolution expressed in the object space cannot be better than $\lambda_c/2$ (coherent case, using a Shannon criterion; the Rayleigh criterion introduces a 1.22 coefficient). This is no longer true in the holographic process when $\lambda \ll \lambda_c$, because the object is not directly imaged at λ_c. To find out what is the true limit, we consider the reconstruction set-up with $z \leqslant y$. This makes the optical magnification $M_O = y/z$ from the source plane to the image plane greater than 1. But the holographic magnification M_H (from object to image) is: $M_H = \rho'/\rho = (z + y)/z$, or $M_H = M_O + 1$. The final resolution can be expressed as the image resolution divided by M_H. In turn, the image resolution is the source resolution times M_O. As, in principle, the source aperture can be as high as 1, the final holographic resolution (apart the aberration problem) can be

$$r_H = (\lambda/2)M_O/M_H$$
$$= (\lambda/2)M_O/(M_O + 1).$$

This shows that the set-up with $z = y$ (optical magnification of 1) produces a holographic magnification of 2, and a best value of the holographic resolution $r_H = \lambda/4$. This result does not violate the law of diffraction. It is due to the fact that information has been gathered into the recorded hologram at λ, which sets a much smaller limit than $\lambda_c/2$, namely $\lambda/2$, for the *recorded* resolution.

4.4.2.3 Correcting the Holographic Aberrations

Reaching this theoretical limit implies dealing with aberrations. Instead of giving a classical analysis of this aberration problem, it is more interesting physically to show how the aberrations work and how the problem can be solved. To this end let us consider the same on-axis point hologram as previously.

According to Sec. 4.3.1 of this Chapter, the information pertinent to the resolution $r = \lambda_c/4$ consists of a spatial frequency $2/\lambda_c$, which is recorded at $\rho = \lambda d/2r = 2\lambda d/\lambda_c$, relative to the center of the point hologram. Having that final resolution requires extraction of the corresponding information; this can be done only if that spatial frequency can diffract an incident ray into a real emergent ray. But considering the law of gratings $\sin\theta + \sin\theta' = \lambda\nu$

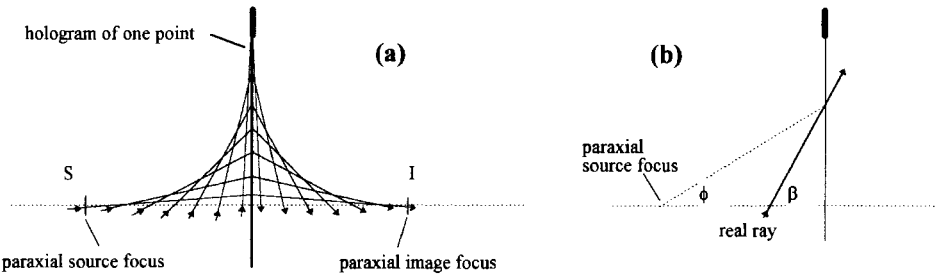


Figure 7. Ray tracing of the aspherical reconstruction wavefront that will diffract the largest possible spatial frequency at λ_c. (a) Such rays are diffracted symmetrically by the hologram. (b) Rays obey $\sin\beta = \tan\phi$ (see text, and Joyeux and Polack [16]).

(ν is the spatial frequency), this is only possible when $\nu = 2/\lambda$, with $\sin\theta = \sin\theta' = 1$, that is, when both incident and emergent rays are at grazing incidence on the hologram. This result illustrates the aberrant behavior of the hologram. But, more important, as the information is located on the hologram at a finite distance, it means that no correction can be obtained if an incident stigmatic wavefront is used, because no ray of a spherical wavefront can reach the hologram plane at a grazing incidence and at a finite distance. For the same reason, no stigmatic (spherical) emergent wavefront can be produced, even if an aspherical incident wavefront is used. In other words, there is no possibility of reconstruction in one step, that is, one incident wavefront producing the corrected image, for the maximum theoretical resolution.

However, this can be done in a two-step process, which happens to be quite simple in practice. The basic idea is first to extract as much information as possible ('primary reconstruction'), which necessarily produces a nonstigmatic image as demonstrated, and then to correct it for its aberrations. The trick is to use the same optics for both operations. This is possible when using the symmetrical reconstruction set-up ($y = z$, $M_H = 2$), provided that a particular aspherical incident wavefront is used. That wavefront is the asphere with a paraxial center of curvature at the paraxial source position, which sends a grazing incident ray to the location of the maximum spatial frequency $2/\lambda$ (Fig. 7), and more generally, which yields a diffracted image wavefront identical to the incident one (apart from the direction of propagation).

Due to this identity, one needs to design only one special optics to produce the desired aspherical incident wavefront from a point source (i.e., from a spherical wavefront), because that same optics, when reversed, also transforms the primary aberrant image into a corrected image. Incidentally, as the magnification of the primary image is only 2, the correction step should also provide an additional large magnification. Finally, using a relief hologram allows the use of a reflected image order. This makes the process even simpler, because one can use the same physical lens forward to illuminate the hologram, and backward to produce the final image. Doing so also eliminates the hologram substrate (a glass plate

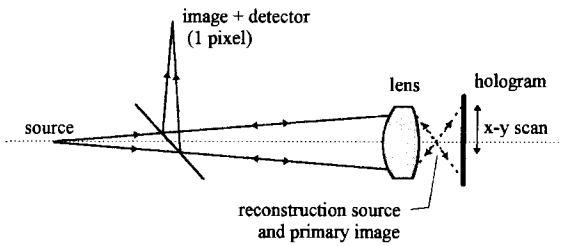


Figure 8. Principle of the scanning reconstruction microscope [15]. The lens-to-hologram distance (working distance) is 300 μm, and the lens focal length is 4.4 mm. Rays on the hologram side are dotted to indicate that the wavefronts are aspherical on that side.

usually) from the optical scheme, avoiding additional aberrations. The resulting device is therefore very similar to a classical microscope (Fig. 8).

So far, we have demonstrated that a single on-axis point hologram can be reconstructed up to a final resolution of $\lambda_c/4$, by using only one specially designed lens. In fact, the field must be accommodated too. It can be demonstrated that the same approach holds, provided that the lens is corrected for the desired field. Undoubtedly, this would lead to a much more complicated optical design, with more optical surfaces. A much better solution is to use a scanning scheme, where the hologram is scanned across the steady reconstruction wavefront, to reconstruct one pixel at a time. This should also solve the optical noise problem, because most of that noise is produced by the diffusion of the optical surfaces: as the image signal is the variation of the flux detected by one pixel-sized detector, any constant parasitic flux is rejected.

This approach has been studied in detail by Joyeux et al. [15]. In principle, the highest possible resolution can be even as high as $\approx\lambda_c/6$, because it is possible to 'immerse' the hologram (as in classical microscopy), which sets the theoretical limit to $\lambda_c/4n$ (where n is the index of refraction; glycerol could be a suitable immersion liquid for use in the ultraviolet, down to $\lambda_c \approx 0.2\,\mu$m). This sets the limit of resolution of optical reconstruction at about 0.033 μm at $\lambda_c \approx 0.2\,\mu$m. A practical limit, when using classical optics technology, is probably close to 0.05 μm. However, the use of SR sources in the visible/ultraviolet, with mirror optics and no immersion may also be considered. It is likely that this latter approach would not give a better practical limit.

Preliminary optical reconstructions have been performed [16], using off the shelf optics, a helium–neon laser, and partial aberration correction. It demonstrated the definite need to get rid of the optical noise. At the present time, a special lens has been designed and manufactured for the reconstruction at $\lambda_c = 0.244\,\mu$m (frequency doubled 0.488 μm argon line) of a hologram recorded with $\lambda d = 7.6\,\mu\text{m}^2$. A scanning 'reconstruction microscope' is being realized, of which the anticipated resolution is 0.1 μm with no immersion. The holographic magnification is 186.

4.5 Fourier Transform Holography

The principle of the Fourier set-up is depicted in Fig. 9. The object can be

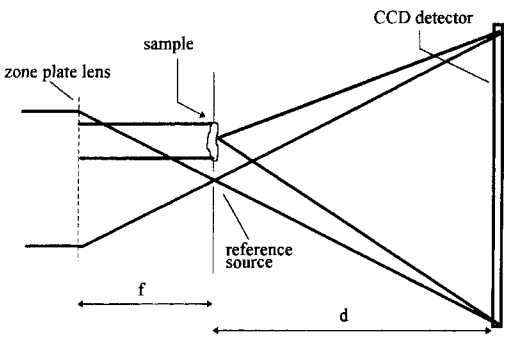


Figure 9. The Fourier transform holographic set-up. The illumination is often provided by the zero (for sample illumination) and focused (for reference) orders of a diffractive lens (Fresnel zone pattern). Typical dimensions are $f = 1$ mm, $d = 80$ mm, and zone plate diameter 100 μm. CCD, charge coupled device. (After McNulty et al. [9].)

illuminated by a more or less plane X-ray wavefront, as in Gabor holography. But a true reference wavefront is added, usually from a point source situated in the object plane, just close to the object field. Compared to Gabor holography, two main differences are noteworthy: (1) the coherence needs and (2) the detector spatial resolution.

Unlike in Gabor holography, the coherence needs are related to the field, because all object pixels are referred to the same reference source. Therefore, it is no longer possible to keep the very large field of Gabor holography, unless a quasi-coherent source is used. However, the flux budget is similar in both cases if the field is kept roughly equal to the coherence zone of the Gabor holography. This is a few tens of micrometers, that is, as long as high resolutions (<0.1 μm) are searched for, a few hundreds of pixels, which is a reasonably large field.

Provided the diffraction angles are small (i.e., resolution ≫ wavelength), the highest spatial frequency to record is $\lambda d/h$, with h the largest distance from reference source to one object point. Thus, the frequency to record can be adjusted through the distance d. In particular, d may be set to large values, so as to make the fringe spacing compatible with the use of an electronic image detector, such as a charge-coupled device, for hologram recording. In this case, the reconstruction should obviously be performed numerically.

The possibility of using an electronic image detector is undoubtedly the main advantage of the Fourier transform configuration. Additional ones are the absence of blurring by other images. Its main drawback is the need for a true reference wavefront. It may be produced by using an X-ray optics having an aperture equal to the diffracting aperture of the object, and focusing it either into a pinhole, or onto a small metallic sphere or a similar diffusing object. But the reference wavefront errors have to be small enough so as not to produce distortion in the reconstruction step, or, alternatively, the reference wavefront must be calibrated, for instance holographically.

This approach was first demonstrated by Reuter et al. [17] under difficult experimental conditions, and then by McNulty et al. [9, 18], who have obtained a 50 nm resolution with test patterns.

4.6 Conclusion

Reading this article, it may have been understood that no holographic instrument presently exists that can routinely perform X-ray holographic microscopy.

However, scientists in the field are still at work, because the method presents specific features that are not shared by other methods: direct phase imaging, from a simultaneous recording of phase and amplitude; an extremely large field; and the ability to act as a simple recording device (Gabor).

Coming back to 3D imaging, we think there is little hope that high-resolution X-ray holograms, obtained as described in this article, could provide a 3D image just as visible light holograms do. This is because the required optical aperture during the recording step seems to be out of our reach for the foreseeable. This does not mean, however, that no 3D imaging can be made using holography, as they certainly can, by using the techniques of aperture synthesis, that is, multiple holograms obtained under different viewing angles. Although also applicable to ordinary imaging, applications to holography may take advantage of the much greater simplicity of the Gabor recording process (as numerous image are necessary), and also of the availability of the recorded phase (during the process of volume reconstruction). To end with, it is worth mentioning that the techniques of *protein crystallography*, used to decipher the 3D structure of proteins from diffraction data obtained with 'reference' diffracting atoms, are conceptually very close to multiple X-ray holography with aperture synthesis. These techniques, which work in the 0.1–0.01 nm spectral range, provide true 3D, subangstrom resolution.

4.7 References

[1] A. V. Baez, *J. Opt. Soc. Am.* **1952,** *42*, 756.

[2] A. V. Baez, H. M. A. El Sum in *X-Ray Microscopy and Microradiography* (Eds.: V. E. Cosslett, A. Engström, H. H. Patte Jr.), Academic, New York **1957,** p. 347.

[3] E. N. Leith, J. Upatnieks, K. A. Haines, *J. Opt. Soc. Am.* **1965,** *55*, 981.

[4] J. W. Giles Jr., *J. Opt. Soc. Am.* **1969,** *59*, 1179.

[5] S. Aoki, Y. Ichihara, S. Kikuta, *Japan J. Appl. Phys.* **1972,** *11*, 1857.

[6] M. Howells, M. Iarocci, J. Kenney, J. Kirz, H. Rarback, *Proc. SPIE* **1984,** *447*, 193.

[7] C. Jacobsen, J. Kirz, M. Howells, K. McQuaid, S. Rothman, R. Feder, D. Sayre in *X-Ray Microscopy II* (Eds.: D. Sayre, M. Howells, J. Kirz, H. Rarback), Springer, Berlin **1988,** p. 253.

[8] D. Joyeux, F. Polack, S. Lowenthal, A. Bernstein in *X-Ray Microscopy II* (Eds.: D. Sayre, M. Howells, J. Kirz, H. Rarback), Springer, Berlin **1988,** p. 246.

[9] I. McNulty, J. Kirz, C. Jacobsen, *Nucl. Instrum. Methods* **1990,** *A291*, 74.

[10] See for instance, M. Born and E. Wolf, *Principles of Optics*, 2nd ed., Pergamon, New York **1964,** p. 511.

[11] G. Koren, F. Polack, D. Joyeux, *J. Opt. Soc. Am(A)* **1993,** *10*, 423, and references therein.

[12] C. Jacobsen, *Ph.D. thesis*, State University of New York at Stony Brook **1988.**

[13] S. Lindaas, *Ph.D. thesis*, Department of Physics, State University of New York **1994.** Also in: S. Lindaas, C. Jacobsen, M. Howells, K. Frank, *Proc. SPIE* **1992,** *1741*, 213.

[14] M. Howells, in *X-Ray Microscopy II* (Eds.: D. Sayre, M. Howells, J. Kirz, H. Rarback), Springer, Berlin **1988,** p. 263.

[15] D. Joyeux, *Proc. SPIE* **1992,** *1741*, 62.

[16] D. Joyeux, F. Polack, in *Short Wavelength Coherent Radiation: Generation and Applications*, Vol. 2, OSA Proc. (Eds.: R. W. Falcone and J. Kirz), OSA, Washington **1988,** p. 295.

[17] B. Reuter, H. Mahr, *J. Phys. E* **1976,** *9*, 746.

[18] I. McNulty, J. E. Trebes, E. H. Anderson, J. M. Brase, W. S. Haddad, D. P. Kern, R. Levesque, H. Szoke Communication at the 4th International Conference on X-ray Microscopy (Chernogolovka, Russia 20–24 Sept. **1993**).

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

5 X-Ray Diffraction Topography

5.1 Introduction

X-ray topography, or X-ray diffraction topography, is the generic name for techniques using X-ray beams that have been Bragg-diffracted by a specimen to image it. The resulting pictures are called topographs. They show the distribution, in direct space as do all microscopy techniques, of various singularities that affect the Bragg reflection used. In particular they show crystal defects such as precipitates, individual dislocations, stacking faults, and domain boundaries.

The principle of the simplest possible version is indicated in Fig. 1. A fairly uniform monochromatic X-ray beam illuminates the single-crystal specimen, set for a Bragg reflection, and an X-ray sensitive photographic plate is placed across the diffracted beam. If the specimen is uniform in all respects, a gray picture without features will be formed corresponding to the shape of the beam, or of the sample if it is bathed entirely by the beam. If, however, the sample contains inhomogeneities that affect the Bragg reflexion used, some areas will receive more X-ray intensity than others, and there will be contrast, hence an image of these inhomogeneities is formed. Geometrically, in this simplest approach, the image corresponds to the projection, along the diffracted beam direction, of the 'defects'. Figure 2 shows two X-ray topographs of a crystal of SiC made by using synchrotron radiation.

Topography is possible because singularities or inhomogeneities can affect the spatial distribution of the diffracted intensity, and hence result in contrast. This is not restricted to X-rays. Indeed neutron diffraction topography is very similar and will be briefly discussed in Sec. 5.7 of this Chapter. In its usual meaning, topography can only be performed on single crystals, or on single grains within a polycrystal. There is a wide range of variants: measurements can be made in transmission or in reflection, with a monochromatic beam or with a white beam, with a divergent beam or with an almost plane wave. Diffraction topography is basically similar to dark-field electron microscopy. It is however very different from electron microscopy in all the relevant orders of magnitude: its resolution is considerably poorer, but it can handle in a completely non-destructive way samples many orders of magnitude thicker.

Diffraction topography is one of the few methods available for making images with radiation for which there are no efficient lenses, and with which neither the classical

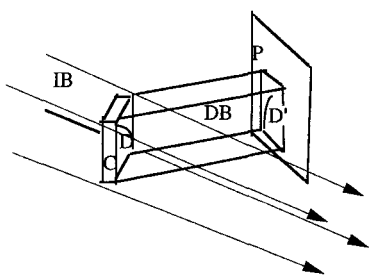


Figure 1. Principle of diffraction topography. IB = incident beam; C = single crystal specimen; DB = diffracted beam; P = photographic plate; D = defect; D′ = image of defect.

approaches of microscopy nor the new scanning probe techniques can be used. Detailed discussions can be found in the literature [1, 2].

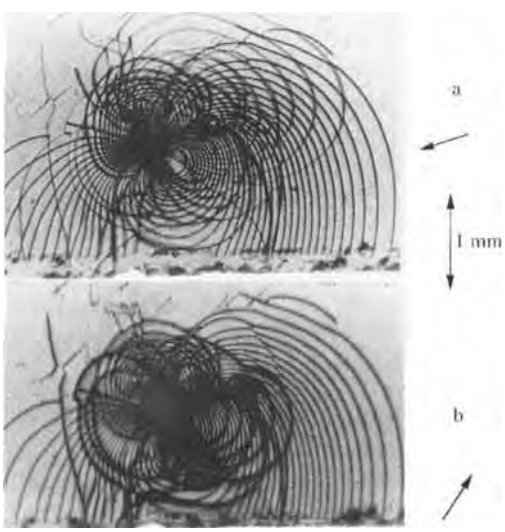


Figure 2. X-ray diffraction topographs of a single crystal (1 0 0) plate of silicon carbide, 300 μm thick, recorded simultaneously using white beam synchrotron radiation topography at ESRF, with exposure time 0.3 s on Kodak type R film. (a) (2 $\bar{1}$ 0) reflection, $\lambda = 0.37$ Å; (b) ($\bar{1}$ 1 $\bar{1}$) reflection, $\lambda = 0.36$ Å. The most conspicuous features are images of single dislocations; some of them are visible in (a) but not in (b). Also note the Pendellösung fringes at the bottom of (a). The arrows indicate the projection of the scattering vector $\boldsymbol{h}$ used. (Courtesy of Eetu Prieur, ESRF.)

5.2 Interaction of X-Rays with Materials (Elastic Processes Only)

We will only discuss hard X-rays, with wavelengths of the order of interatomic distances, (i.e., $\sim 10^{-10}$ m). The incentive for using X-rays for imaging is that their interaction with materials is quite different, both qualitatively and quantitatively, from that of the usual probes of microscopy, namely visible light, sound, and electrons. Whereas visible light is affected by the weakly bound electron states, sound by the elastic properties, and electrons by the electrostatic potential, X-rays are scattered by the electron density at large, in which inner shells play an important part. Unlike light and sound, where the wavelengths are large compared to the interatomic distances, X-rays are sensitive to the long-range but short-period spatial order that characterizes crystals and quasi-crystals, and produces Bragg diffraction spots. This is the basis for the use of X-rays in structural crystallography, for determining the shape and dimensions of the unit cell and for characterizing the positions of the atoms and their thermal vibrations. Most of the observations in X-ray topography show the local deviations from perfect periodicity, or the local changes in periodicity, associated with in defects. In comparison with electrons, which share this sensitivity to long-range order, and in particular to lattice periodicity, X-rays stand out through the fact that their interaction with materials is much weaker, in terms both of absorption and of scattering, and that, therefore, much thicker samples can and must be used. Since the absence of lenses makes it

impossible to obtain a high magnification, the investigated areas are also much larger. Hence, X-ray topography is a technique for investigating bulk samples.

The absorption of X-rays is almost always an atomic process in the sense that all atoms contribute regardless of their arrangement. The basic mechanism, formalism and orders of magnitude are discussed in this Volume by Mouze (see Chap. 2, Sec. 2 of this Volume). The exception is the case of Bragg diffraction in perfect crystals: as discussed below, anomalous transmission (i.e., the Borrmann effect), then sets in.

Structural crystallography is based on the use of the kinematical approximation for the description of elastic scattering. It consists in assuming that the incident wave reaching all points in the specimen has the same amplitude, and is just shifted in phase. This approach will be assumed to be familiar to the reader, and can be found in classical crystallography textbooks [3, 4]. Only the results of the method will be summarized here. Consider a small sample around origin O submitted to an incident plane wave with amplitude E_O and wave vector $\boldsymbol{k}_O$. The scattered wave induced at point $\boldsymbol{r}$, at large distance from origin O, can be written as

$$E_O \frac{f(\boldsymbol{q})}{r} \exp(2\pi i k r) \tag{1}$$

with $k = k_O = 1/\lambda$ and $\boldsymbol{q} = \boldsymbol{k} - \boldsymbol{k}_O$ the scattering vector, with $\boldsymbol{k}$ along $\boldsymbol{r}$. The scattering length, $f(\boldsymbol{q})$, becomes

$$f(\boldsymbol{q}) = -r_c \cdot C_i \tilde{\rho}(\boldsymbol{q}) \tag{2}$$

with $r_c = e^2/4\pi\epsilon_0 mc^2 = 2.8 \times 10^{-15}$ m, or 2.8 fm (femtometer or fermi) traditionally called the classical electron radius and $\tilde{\rho}(\boldsymbol{q})$ the Fourier transform of the electron density in the sample. The polarization coefficient C_i, with i designating the polarization state, results from the vector nature of X-rays: $C_\sigma = 1$ for the incident electric field perpendicular to the scattering plane $(\boldsymbol{k}_O, \boldsymbol{k})$, and $C_\pi = \cos 2\theta$ when it is in this plane. An unpolarized incident beam results in the diffracted intensity involving $(1 + \cos^2 2\theta)/2$.

The amplitude scattered by an atom at origin is $-f_{at}^0(q) r_c$ if the electrons are regarded as free. The atomic scattering factor, $f_{at}^0(q)$, a number in X-ray usage, is the Fourier transform of the electron density in the atom. This is tabulated in [5], as a function of $(\sin\theta)/\lambda = q/2$. It has a monotonic decrease with increasing q, and increases regularly with increasing atomic number Z. Its value for $q = 0$ is indeed just Z. A correction has to be made when the X-ray energy is near one of the electron binding energies, that is, near an absorption edge. Then energy-dependent dispersion corrections set in, and $f_{at} = f_{at}^0 + f'_{at} + i f''_{at}$, with f' and f'' real.

In so far as the kinematical approximation can be considered valid, the result extends to a small crystal. Then the electron density is almost periodic in space, hence its Fourier transform is a Fourier series, with each term just broadened by the fact that the object is limited in space. The only $\boldsymbol{q}$ values for which there is diffracted intensity are very near the reciprocal-lattice vectors,

$$\boldsymbol{h} = h\boldsymbol{a}^* + k\boldsymbol{b}^* + l\boldsymbol{c}^* \tag{3}$$

The condition for diffraction $\boldsymbol{q} = \boldsymbol{k} - \boldsymbol{k}_O = \boldsymbol{h}$ is equivalent to the condition for constructive interference from the scatterers around each lattice node. It can easily be put into the familiar Bragg-reflection form: $2d_{hkl} \sin\theta = \lambda$, where d_{hkl}

is the spacing between neighboring (hkl) lattice planes and θ the angle of the incident and diffracted beams with these planes. The diffracted wave can be approximated at very large distances as a plane wave with wave-vector $\boldsymbol{k} = \boldsymbol{k}_O + \boldsymbol{h}$. Its amplitude is proportional to the coefficient $\rho_{\boldsymbol{h}}$ in the Fourier series expansion of the electron density. The structure factor $F_{\boldsymbol{h}}$ describes the amplitude diffracted by one unit cell at origin. It is easily shown that $F_{\boldsymbol{h}} = \rho_{\boldsymbol{h}} V_c$, where V_c is the volume of the unit cell, can be expressed, in terms of the atomic scattering factors $f_{at\,i}$ of the atoms at coordinates $\boldsymbol{r}_i$, as

$$F_{\boldsymbol{h}} = \sum_i f_{at\,i} \exp(2\pi i \boldsymbol{h} \cdot \boldsymbol{r}_i) \cdot \exp(-W_i), \quad (4)$$

where the last term, the Debye–Waller factor, accounts for the thermal vibration of the atoms. Thus $\rho_{\boldsymbol{h}}$ or $F_{\boldsymbol{h}}/V_c$ is the weight to be associated with a reciprocal lattice node $\boldsymbol{h}$, and the larger $|F_{\boldsymbol{h}}|$, the stronger is reflection $\boldsymbol{h}$ or hkl. Quantitatively, the intensity of a reflection is expressed in terms, not of the intensity obtained in the unrealistic case of an incident plane wave, but as the area N_{int} under the rocking-curve, that is, the curve giving the diffracted intensity as a function of angle ω or time when the crystal is rotated around the orientation where Bragg reflection $\boldsymbol{h}$ occurs. This is proportional to the incident intensity I_O, and inversely proportional to the rocking speed $\dot{\omega}$:

$$N_{int} = r_{\boldsymbol{h}} \frac{I_O}{\dot{\omega}} \quad (5)$$

In the simplest experimental arrangement, the expression for the integrated reflectivity of reflection $\boldsymbol{h}$ is

$$r_{\boldsymbol{h}} = r_c^2 \lambda^3 \frac{1 + \cos^2 2\theta}{2} \frac{1}{\sin 2\theta} \frac{|F_{\boldsymbol{h}}|^2}{V_c^2} v \quad (6)$$

where v is the volume of the specimen, under the assumption of unpolarized incident radiation bathing the crystal entirely.

The assumption that the incident beam has the same amplitude everywhere in the crystal, the basis for the kinematical approximation, contradicts energy conservation. Energy is transferred from the incident beam into the diffracted beam(s), and the kinematical approximation is not tenable if the crystal is large and the reflection strong. The dynamical theory of diffraction takes into account the mutual exchange between incident and diffracted beam(s). It contains a wealth of detailed predictions on the diffraction behavior of perfect crystals, and is discussed in details in [6, 7]. Here we simply summarize, without proof, the four points that are relevant to topography, in the case where only the incident and one diffracted beam are present (2-beam case). The first point sets the order of magnitude of the reflectivity of perfect crystals for the simplest situation, that of a monochromatic beam. In reflection (Bragg) geometry (Fig. 3), for a thick crystal, if absorption is neglected, total reflection occurs (i.e., reflectivity is 1), over an angular range

$$\epsilon = \frac{r_c \lambda^2 |F_{\boldsymbol{h}}|}{\pi V_c \sin 2\theta} = \frac{d_{hkl}}{\Lambda} \quad (7)$$

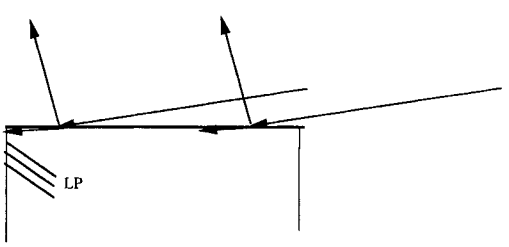


Figure 3. Dynamical diffraction, Bragg (reflection) geometry. The lattice planes used for the Bragg reflection are indicated as LP.

ϵ is the Darwin width. As a typical order of magnitude, the values for reflection (2 2 0) from silicon, for $Cu_{K\alpha}$ radiation are $\Lambda = 16\,\mu m$, $\epsilon = 2.5$ arcsec. The integrated reflectivity is then simply

$$r'_h = \frac{\pi\epsilon}{2} \approx 2 \times 10^{-5} \tag{8}$$

This reflectivity is different from that used in the kinematical approximation because r'_h describes the situation of an extended crystal, where the cross-section of the diffracted beam is not determined by the crystal. Accordingly r'_h is dimensionally a number instead of an area. The second point is the interference effect called Pendellösung. The incident beam and the diffracted beam are comparable to two coupled pendulums: it is a familiar observation that, if one of the pendulums is excited initially, it will after some time slow down, while the other one takes on increasing oscillation amplitude, until the exchange reverses etc. This behavior can be carried over to the two beams: their coupling occurs in space via the crystal, and it is characterized by the strength of the reflection (i.e., by F_h/V_c). When a plane wave is incident on the crystal, in transmission (Laue) geometry, the incident (or, more exactly, very slightly refracted) wave will start out alone near the entrance surface. It will progressively decrease in amplitude while creating the diffracted wave, which will then subside and leave place for a growing refracted wave etc. (Fig. 4). The refracted and diffracted wave, coupled by the crystal, make up an entity called a wave-field. On leaving the crystal, they separate out into two beams, one in the direction $\boldsymbol{k}_O$ or O of the incident beam, the other in the diffracted direction, along $\boldsymbol{k} = \boldsymbol{k}_O + \boldsymbol{h}$. The interference, or

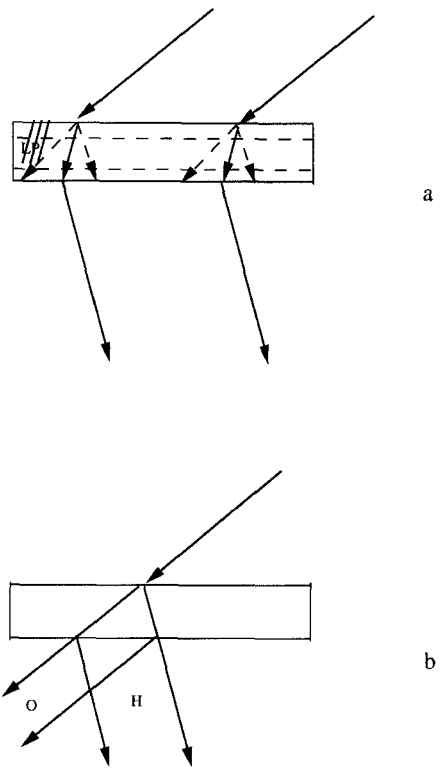


Figure 4. Dynamical diffraction, Laue (transmission) geometry. (a) For an incident plane wave, the horizontal dotted lines could correspond to loci of maximum diffracted intensity, they would be spaced Λ apart for the simplest case of symmetrical diffraction (the diffracting lattice planes, LP, would then be perpendicular to the entrance face of the sample) and for an incident wave exactly satisfying Bragg's conditions; (b) for an incident wave severely restricted in width, hence divergent (cylindrical wave).

Pendellösung, fringes associated to the oscillation in intensity are called Pendellösung fringes. The characteristic length for the exchange, the Pendellösung period, is, in the simplest case:

$$\Lambda = \frac{\pi V_c \cos\theta}{r_c \lambda |F_h|} \tag{9}$$

The third important result of dynamical theory is the behavior of integrated reflectivity against crystal thickness t. The

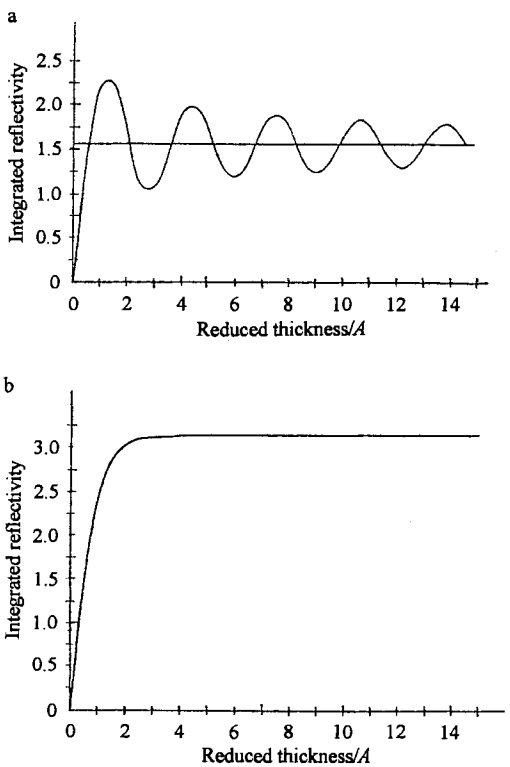


Figure 5. Integrated reflectivity, scaled with $\epsilon/2$, vs crystal thickness, scaled with Λ/π, for Bragg diffraction by a perfect crystal, in zero absorption limit: (a) Laue geometry; (b) Bragg geometry. Kinematical behavior corresponds to the continuation of the initial slope.

results are best expressed as the scaled integrated reflectivity $(2r'_h)/\epsilon$ in terms of the reduced variable $A = \pi t/\Lambda$, and they are shown in Fig. 5, for both the Bragg (reflection) and Laue (transmission) geometries, in the zero absorption limit. The kinematical approximation corresponds to the initial slope of either of the curves, implying the proportionality to crystal volume which stands out in the expression of the kinematical r_h. The very important feature is that the perfect crystal reflectivity as given by the dynamical theory is smaller than the kinematical approximation, and tends towards this value only for small crystal thickness $t \ll \Lambda$. This decrease in diffracted intensity with respect to the kinematical approximation when the crystal is too good is called extinction. Since Λ involves the structure factor, we confirm the qualitative discussion above: the kinematical approximation is valid for small t or a weak reflection. A more difficult argument shows that the kinematical approximation remains valid for larger but imperfect crystals, that is, crystals featuring distortions where Bragg's law is not exactly satisfied at all points for a given wave.

The fourth spectacular result concerns absorption. In a perfect crystal, interference between the refracted and the diffracted components of a wave-field lead to a stationary component with equal intensity loci parallel to the lattice planes that perform the Bragg 'reflection'. For one of the wavefields, the nodes of the standing wave are on the atoms. Then the X-ray intensity is very low at the positions where absorption occurs, namely, in the internal electron shells of the atoms, and accordingly absorption is reduced in comparison with the simple standard approach. Absorption is enhanced, because the antinodes are on the atom planes, for the other wavefield. The Borrmann effect is the spectacular physical consequence: very absorbing crystals ($\mu t > 10$) can transmit significant intensity if they are perfect and when exactly set for Bragg diffraction. In this case two beams with equal intensities leave the crystal: the transmitted (actually forward-diffracted) and the diffracted beam.

The effect of isolated defects, with a known spatial distribution of lattice distortion, on the propagation of wavefields in a perfect crystal matrix is well

understood, qualitatively, numerically and sometimes analytically. This is the basis for the description of the contrast produced by defects on X-ray topographs. The computer modeling of contrast is based on the Takagi–Taupin equations for the propagation of wavefields in the presence of lattice distortions. The simulation of topographs is now a very active and efficient area [8].

5.3 Principle of Bragg Diffraction Imaging

It should now be at least plausible that inhomogeneity in diffracted intensity, that is, contrast on the topographs, can arise through several mechanisms [9]. A change in the orientation of the reflecting planes, hence of the reciprocal lattice vector $\boldsymbol{h}$, means a change in the geometrical condition for Bragg reflection. Contrast related to this 'misorientation' effect can take different forms depending on the X-ray beam and geometry used. A region with a rapid spatial variation of distortion (gradient of effective distortion $>\epsilon/\Lambda$) looks to the X-rays as a piece of imperfect crystal, corresponding to the kinematical approximation. It will therefore diffract more intensity, if the absorption is low, than the perfect crystal matrix, and be imaged as a 'direct' or 'kinematical' image, with 'extinction contrast' corresponding to more intensity locally. This is the case in Fig. 2. However, the reverse will be true in a situation of high absorption ($\mu t \gg 1$), because the imperfect part will locally disrupt the Borrmann effect and result in lower local intensity. Slow spatial variations in distortion give rise to changes in the propagation of wavefields, somewhat analogous to mirages in optics, resulting in the 'dynamical' image. Contrast also arises without crystal defects, in low-absorption cases, if the sample thickness is inhomogeneous, because the diffracted intensity varies with thickness. It takes the form of Pendellösung fringes corresponding to equal-thickness contours, spaced Λ apart, as can be seen in the bottom part of Fig. 2a. In the presence of a crystal defect, new interference fringe systems involving wavefields that have been deflected by the defect can arise; they make up the 'intermediate' image.

The width of the image of isolated defects (e.g., dislocation lines), is determined by the various ways they affect the propagation of wavefields in the crystal to give rise to the kinematical, intermediate and dynamical images [9]. In the case of the kinematical image, this width is approximately the distance from the defect core where the misorientation of the lattice planes used is equal to the Darwin width ϵ. Because ϵ is much smaller than in electron diffraction due to the fact that the interaction of X-rays with the crystal, as defined in terms of a scattering length by $r_c|F_{\boldsymbol{h}}|$, is much smaller, the intrinsic width of the dislocation images, 1–10 μm, is much broader than in electron microscopy. This is a fortunate circumstance, since it gives the possibility to see individual dislocations although there is no magnification in Lang's method. But it implies that the resolution limit, typically a couple of micrometers at best, is intrinsically far less good than in electron microscopy. Magnifications of about 100 are usually quite sufficient for X-ray

topographs, and the standard values are much less.

The spatial resolution is better when the topographic technique is sensitive only to large distortions, because these are nearer the defect core. It therefore varies inversely with the angular resolution. High angular resolution can be achieved by double-crystal methods in which the monochromator geometry is selected to make the crystal rocking-curve very narrow. Very small local deviations of the lattice planes or changes in their distance then leads to high contrast [1]. At the other limit, the weak-beam technique, derived from electron microscopy, consists in setting the sample far from Bragg's condition for the perfect crystal matrix, so that only highly distorted parts contribute to the image, which is then very narrow.

The local variations in intensity in the diffracted beam(s) that are the basis for topography must be recorded. The usual recording medium is X-ray film, or nuclear emulsion plates. Film is more convenient and cheaper. It is manufactured for industrial or medical radiography purposes, and comes in a variety of grades (i.e., of sensitivity) and grain sizes. At topography stations with a large throughput (i.e., at synchrotron radiation sources, see below), films can be developed by commercial machines. Nuclear emulsion plates have the disadvantages of high cost and tedious processing; they have a thick coating that makes better use of the X-ray intensity because it absorbs most of it, they are very fine-grained, and the glass substrate is considerably more resistant than the plastic base of films. The main supplier is Ilford Company, Ilford (UK). Typical exposure times are of the order of 10 h on fine-focus sealed-tube generators, 1 h on rotating anode generators, and seconds or less on synchrotron radiation sources. Electronic detectors are at present not as good in resolution as films or nuclear emulsion plates, but they are invaluable for the observation in real time of phase transitions, domain evolution etc. Several versions exist, and work well on high intensity sources, viz. rotating anode generators, or synchrotron radiation sources. Their design is based on a scintillator combined with a low light level television camera. On comparatively high divergence beams, that is, in all cases except for synchrotron radiation, the detector must be placed as close to the specimen as possible. When topography is to be performed at low temperatures, and/or under a magnetic field, this implies a special design for cryostat tails and electromagnet.

Standard topography is based on the diffracted beam(s) because, in most cases, the transmitted beam is only weakly affected by the diffraction processes and would have very low contrast. However the forward-diffracted beam can be used just as well as the diffracted one in the case of high absorption, since their joint survival is due to the same diffraction process and is disrupted by imperfections in the same way, while the simply transmitted beam is then effectively absorbed out. In this case the film can even be in contact with the specimen.

5.4 X-Ray Sources and Beams

There are two types of sources of X-rays: laboratory generators, and synchrotron

radiation facilities [10]. Modern laboratory X-ray generators are either of the sealed tube, or of the rotating anode type. The former are cheaper and simple to maintain, the latter have very much higher brilliance. In either case, the principle of X-ray production is to focus a beam of high energy electrons on a small area of a metal anode cooled by water circulation. A variety of inelastic collisions give rise to accelerations of the electrons which give rise, through the same mechanism as discussed in the framework of scattering above, to photons forming a continuous spectrum called Bremsstrahlung. The characteristic spectrum is excited provided the electrons' energy is higher than the binding energy of deep levels, and therefore than the K, L_1, L_2, L_3 ... absorption edges. The relaxation following the production of a hole on a K shell involves sharp transitions, hence narrow photon emission lines, such as K–L_2 ($K\alpha_2$ line), K–L_3 ($K\alpha_1$ line), K–M_3 ($K\beta_1$), etc., as well as L emission lines. The wavelengths of the emission lines and of the absorption edges are tabulated in the literature [10]. Values for copper are: $K\alpha_1$: 1.54051 Å; $K\alpha_2$: 1.54433 Å; $K\beta_2$: 1.39217 Å; K absorption edge at 1.380 Å. The intensity of $K\alpha_1$ component of the $K\alpha$ doublet being the highest (about twice that of the $K\alpha_2$), it is the favorite radiation in X-ray topography from generators. Rotating anode generators can provide a very high brilliance because the focal spot, while remaining fixed with respect to the laboratory, is thermally spread out over the circumference of the anode.

Synchrotron radiation is also based on radiation by an accelerated charge, but the essential item is that in this case the charges, circulating on a curved orbit, are highly relativistic, that is, their velocity is extremely close to that of light [10, 11]. As a result, their radiation pattern, which is symmetric in a reference frame in which the particles are at rest, appears from the laboratory point of view as highly distorted, with a peak in the forward direction with angular width $1/\gamma$ where $\gamma = E/(mc^2)$ is the ratio of the particles' energy to rest mass. In the practical case of electrons or positrons, $mc^2 = 0.511$ MeV, while E is (for example) 6 GeV at the European Synchrotron Radiation Facility in Grenoble, France (ESRF). As a result, a point in the laboratory is reached by the beam from a bunch of electrons for an extremely short time, and therefore, through Fourier transformation, the energy spectrum is very broad. In fact, this 'white' spectrum spans the range from infrared to hard X-rays in such a machine. The simplest type of synchrotron radiation arrangement involves bending magnets: the curvature of the electron orbit is produced by applying a uniform magnetic field. In modern synchrotron radiation sources, the beam spectrum can be tailored to the need of an experiment through insertion devices, placed in the straight sections of the electron orbits. These devices create non-uniform magnetic fields, and entail a noncircular portion of orbit. Among them are wavelength shifters, wigglers and undulators, giving a large degree of flexibility as well as unsurpassed brilliance (Fig. 6). Other potentially useful characteristics are the polarization of the beam, its coherence, and its time structure. The polarization corresponds, in the plane of the electron orbit, to $\boldsymbol{E}$ being parallel to this (horizontal) plane because, except in very special insertion devices, the acceleration of the electrons is horizontal

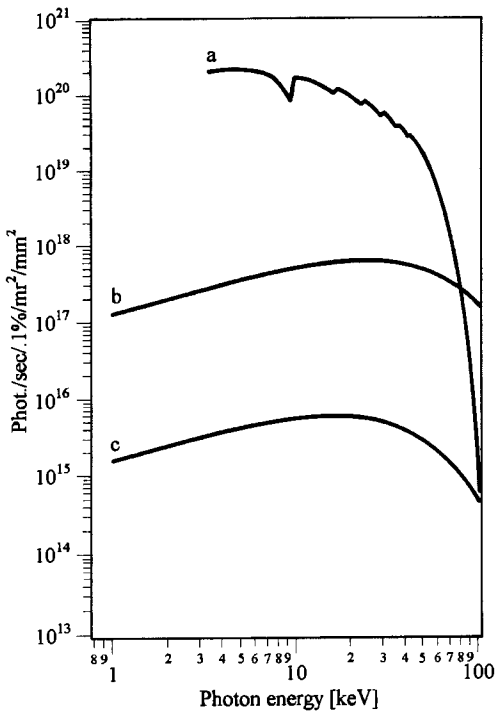


Figure 6. Brilliance of synchrotron radiation at ESRF, for the electron beam current of 200 mA which will become standard. (Courtesy of Pascal Elleaume, ESRF.) Curve a, 5 m undulator; curve b, 1.25 T wiggler; curve c, 0.85 T bending magnet. The brilliance is expressed as a function of the photon energy as the number of photons emitted per second, in a spectral bandwidth of 0.1%, per $(\text{mrad})^2$ of solid angle, per mm^2 of source area.

too. Above and below the plane of the orbit, the polarization changes to elliptical, while the intensity decreases rapidly. The pulsed time structure of the electron beam, a necessity for machine operation, gives a pulsed structure to the photon beams too.

In all cases, the X-ray beams to be used for X-ray topography have to be conditioned to some extent. The restriction of the beam cross-section to the appropriate size, through a slit, is always performed through absorption by high atomic weight material (lead or tungsten carbide). It can be less simple than it appears, either when the beam is to be very fine (see Sec. 5.5 of this Chapter), or when the beam is very intense and thus carries a lot of power (white beam of synchrotron radiation). In fact stopping the small amount of enormously energetic γ-rays produced in the synchrotron by collision of electrons with the residual atoms in the ring is very difficult. Monochromatization involves Bragg reflection from a monochromator (i.e., a single crystal), often silicon because very perfect crystals are readily available and the technology for machining it is very well established, with the Bragg reflection appropriately chosen. The use of a perfect crystal monochromator provides a lot of flexibility in terms of the shape of the beam delivered to the sample and of the conditions for scattering [1]. The result can be an almost plane wave with a broad cross section, or a fairly divergent beam with a small cross-section, and various types of variation of wavelength with angle (i.e., dispersion) characteristics. The angular collimation of the incident beam, whether monochromatized or not, can be defined as the divergence of the beam reaching a point in the specimen. Synchrotron radiation X-rays usually have quite a small divergence because the effective source, viz. the part of the electron or positron beam seen from the sample position, is small and far from the sample: thus at a bending magnet on the ESRF, the source size is about 0.5 mm × 0.35 mm, at a distance of 60 m, resulting in a vertical divergence of 6 μrad and a horizontal divergence of 8 μrad. With laboratory X-rays, small focus sealed tubes, or fine focus settings on rotating anode generators, have to be used.

5.5 Implementation of X-Ray Topography

More detailed descriptions of the different methods and their possibilities are available [1, 2].

X-ray topography started on laboratory X-ray generators, and most of the topographic work is still based on their use. The presence of characteristic lines obviates in most cases the need to monochromatize the radiation.

The oldest among the modern techniques, the Berg–Barrett method of reflection topography, probes the surface of the sample. It shows the subgrains, and gives the possibility of measuring their misorientation.

Transmission topography on laboratory generators was given its modern form by Lang in 1957 and 1958, when he invented the two forms known as section topography, and traverse topography (Fig. 7). The basic idea behind both is to obviate the doubling of images that the doublet nature of the Kα characteristic emission line would normally give. This is achieved by using a fine focus generator, and restricting the beam incident on the specimen to a slit, narrow enough that the divergence be smaller than the difference in Bragg angles for Kα_1 and Kα_2. In the case of section topography, the beam is typically 10 to 20 µm broad. When, as a first approximation, only kinematical images are considered, then only the defects contained in the virtual slice defined by the incident beam are imaged, whence the name section topography. It then becomes desirable to investigate a larger volume of the sample, but without opening up the slit very much (100 to 200 µm) lest the images appear double. The trick, invented by Lang, was to traverse simultaneously the sample and the photographic detector, while leaving fixed relative to the incident beam a slit shielding the film from the transmitted beam. This slit can be used to select kinematical images from only part of the crystal thickness. Lang cameras capable of performing both forms of topography are available from several manufacturers. They are precision instruments both because of the slit and because of the traverse mechanism. Modern forms include computer help for measuring rocking-curves, setting the specimen for exposure at the top or on the sides of the rocking-curve, etc., and accept very large samples as produced by the microelectronics industry. From the basic point of view, a traverse topograph is the superposition of section topographs made at neighboring positions. Section topographs can be understood as a projection of the distribution of diffracted intensity across

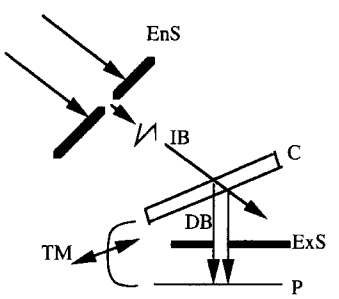


Figure 7. Lang's methods of section and traverse topography: IB = incident beam; EnS = entrance slit; C = single crystal specimen; DB = diffracted beam; ExS = exit slit; P = photographic plate; TM = traverse movement, simultaneously moving the crystal and the film for traverse topographs. For section topographs, TM is not activated, and the entrance slit is narrower (10 to 20 µm) than for traverse topographs.

the exit surface of the crystal when the incident wave is restricted to a fine slit and is therefore a diverging cylindrical wave. The diffracted intensity at any point of the exit surface is the integrated intensity corresponding to the addition of contributions from plane-wave components that have slightly different wave-vectors at the entrance, and entail very different wave-field paths in the crystal. Section topographs are of considerable assistance in understanding the contrast when the first approximation referred to above, viz. that of kinematical images only, is dropped.

Synchrotron radiation can be used for X-ray topography in a variety of forms. In the white-beam technique, the incident white beam directly illuminates the stationary sample. The Laue diagram thus produced consists of several diffraction spots, namely all those satisfying Bragg's relation for the angle θ between the beam and a set of lattice planes. Because the beam divergence is very small, each spot is a topograph with a decent resolution, even if the film is placed at distances of the order of 10 cm or more. More information, including the invisibility of some defects, giving indications on their characteristics, can thus be obtained in one exposure, although the variation in density among the spots may require different exposure times. Figure 2 shows a few of the topographs thus obtained in one shot. Setting an X-ray sensitive camera across one of the spots provides a simple way of following the real-time evolution of the image under the effect of changes in temperature or applied fields. Synchrotron radiation topography can also be performed with a monochromator, and the very high intensity is particularly useful for

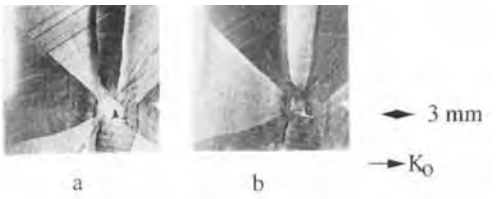


Figure 8. Almost plane-wave reflection topographs of a $(1\,\bar{1}\,0)$ plate of $Sr(NO_3)_2$, taken at the synchrotron radiation source LURE, showing growth sectors. Reflection $(8\,\bar{8}\,0)$, $\lambda = 1.24$ Å. The projection of the incident beam direction $\boldsymbol{K}_O$ is indicated by the arrow. The two topographs in (a) and (b) correspond to situations in which the crystal orientation was changed by 2″. (Courtesy of Marie-Claire Robert, Laboratoire de Minéralogie et Cristallographie, Universités de Paris 6 et 7.)

performing extreme versions, in particular quasi-plane wave topography. An example of topographs made using this form of topography, also called double-crystal topography, is shown on Fig. 8. It can feature very high sensitivity to differences in orientation between parts of a crystal, for example, between different growth sectors for crystals grown in solution.

5.6 Possibilities of X-Ray Topography

One basic limitation of X-ray topography must be emphasized again: it can only image single crystals, or individual grains within a polycrystal. Although the influence of surface treatment can be of interest, it usually aims at giving information about crystal defects and/or domains, and it is then desirable to eliminate strained surface regions through chemical dissolution or electrochemical polishing. X-ray topography provides information on the lattice distortion, that is, on changes in the

orientation and modulus of the reciprocal lattice vector $\boldsymbol{h}$ or, equivalently, on the direction and spacing of the lattice planes $(h\,k\,l)$ used for the reflection. The important quantity is the effective change in Bragg's angle for a given reflexion $\boldsymbol{h}$ that is associated with the distortion: as shown in [8], this can be expressed as a function of the displacement $\boldsymbol{u}$ by

$$\delta\theta = -\frac{\lambda}{\sin 2\theta}\,\frac{\partial(\mathbf{h}\cdot\mathbf{u})}{\partial x_h} \qquad (10)$$

where x_h is a coordinate along the diffracted beam direction. As a result, topography can show individual crystal defects, namely, inclusions, dislocations, stacking faults, twins. Simple refinements of the techniques make it possible to get a very spectacular stereographic view of the defects [1]. Just as dark-field electron microscopy, topography can also characterize the defects, because there is no distortion for some lattice planes. For example, screw dislocations are invisible when diffraction is performed from planes containing the dislocation line and Burgers vector $\boldsymbol{b}$ (i.e., when $\mathbf{h}\cdot\mathbf{b} = 0$). The density of defects that can be resolved is limited by the condition that the parts of their strain fields effective for the reflection used do not overlap. The simple case of screw dislocations gives an image width $(\Lambda/2\pi)\mathbf{h}\cdot\mathbf{b}$ [2]. As a result, the maximum density of dislocations is $\approx 10^6\,\mathrm{cm}^{-2}$ using standard topography techniques. The same mechanism, namely the change in diffracted intensity due to inhomogeneous lattice distortion, also makes it possible to see growth sectors, growth striation, and various kinds of domains. While ferroelastic domains differ in their distortion by definition, there are also differences in distortion in most cases of ferroelectric domains, and, through magnetostriction, for non-180° ferromagnetic domains. Different lattice distortion can occur between antiferromagnetic domains through exchange striction as well as magnetostriction, although in many cases classical X-ray or neutron structural diffraction techniques are not sensitive enough to detect the distortion. First order phase transitions in the solid phase necessarily involve coexistence of phases with different lattice distortions, and here again X-ray topography is very well suited for observing the interphase morphology and the evolution under a change in temperature or field. Spectacular effects of inhomogeneous increase in diffracted intensity have been observed in many insulators or ionic conductors (e.g., quartz, or α-lithium iodate) when submitted to an electric field. They are not well understood at the present time. In the technologically important systems involving epitactic layers that are manufactured for electronic and optoelectronic applications, beautiful and informative moiré fringes, as shown in Fig. 9, coexist with images of dislocations in the substrate and of misfit dislocation at an interface. In such samples where a layer of silicon is grown on silicon oxide produced on the silicon substrate, these moiré fringes occur naturally, in contrast to the difficult techniques which Lang pioneered for their observation in separate crystals [1], and show the deviation in orientation between the substrate and silicon layer.

5.7 Neutron Topography

Neutron diffraction topography is worth mentioning here because it is similar and

Figure 9. Synchrotron radiation topograph of silicon on insulator (Simox) sample, consisting of (1 0 0) silicon substrate, 550 μm thick, with 0.4 μm layer of SiO_2 and 15 μm of Si epitaxically grown on top. (0 4 0) reflection, $\lambda = 0.3$ Å, exposure time 1000 s on HR film. Dislocations in the substrate, misfit dislocations, and moiré fringes are visible. (Courtesy of Eetu Prieur, ESRF.)

complementary, although it is less developed than its X-ray counterpart, among others because there are no laboratory sources of neutrons. Neutrons in the energy range around 10^{-2} eV (i.e., those designated as hot, thermal and cold) have a de Broglie wavelength in the 0.1–10 Å range, and behave very much like X-rays in the geometrical features of their diffraction by crystals [12]. When only nuclear scattering is involved, both the kinematical approximation and the dynamical theory can be directly transferred from the X-ray case by substituting b_c, the coherent scattering length, as tabulated in [13], for $(r_c f_{at}(q))$. The coherent scattering lengths are q-independent because the strong-force interaction responsible for the neutron-nucleus interaction is very short-ranged, and only a little smaller than the $r_c f_{at}(q)$ valid for X-rays. The fact that they have no systematic variation with atomic number is the basis for the use of neutrons in structural crystallography, where they are very valuable in locating light atoms such as deuterium, in the presence of heavy ones, or distinguishing between atoms with close-by atomic numbers. Neutron topography is obviously possible, and neutrons are sensitive to lattice distortion, hence they can be used in complete analogy with X-ray topography. However, the resolution of neutron topography is poor (typically about 100 μm) essentially because there are few neutrons even in beams of a high flux reactor [14]. It is therefore not well suited to investigating individual defects, although their observation was shown to be possible. Two features make neutrons very valuable for topographic purposes. One is the fact that the absorption of neutrons by most materials is very weak, and here again not related to atomic number. Samples containing heavy elements can be very transparent to neutrons at thicknesses where their absorption for X-rays would be very high. Thus the investigation of as-grown crystals, natural or artificial, becomes possible in transmission even for high Z materials. In particular, the method of section topography as described in Sec. 5.5 of this Chapter acquires a new dimension. The other important feature is that neutrons have a magnetic moment, and hence are sensitive to the distribution of magnetic moments in a crystal. This is the basis for the use of neutrons in determining magnetic structures, that is, the arrangement of (electronic) atomic moments within the unit cell. In the topographic approach, the interest shifts to the inhomogeneity in magnetic structure, that is, in magnetic domains of all kinds. In this case, the feature of interest is the local change of the magnetic contribution to the structure factor F_h, resulting in different reflectivities from different domains, in particular with polarized neutrons.

Whereas there are several much more convenient methods for observing ferro- and ferrimagnetic domains, neutron topography is the only way of observing antiferromagnetic domains directly, in the sense that this observation is based on the definition, not on a side effect such as striction. It is in fact the only method for visualizing some types of domains, for example, chirality domains in helimagnets (right/left-hand screw structure) such as terbium, and 180° or time-reversed domains in MnF_2. Magnetic phase transitions involving the coexistence of different magnetic structures can also be investigated very directly by neutron topography [15].

5.8 Conclusion

X-ray topography can be performed in the laboratory or at synchrotron radiation sources. It is non-destructive, and very well suited both to comparison with other observation techniques and to the application of variable temperature and applied fields or stress. In comparison with electron and optical microscopy, its most significant features are the high sensitivity to strain, and the ability to handle thick (10^{-4}–10^{-3} m) samples of non-optically transparent materials. The resolution can be tailored to a large extent, with the spatial resolution $\geqslant 10^{-6}$ m and the angular resolution $\geqslant 10^{-7}$ rad. Synchrotron radiation provides powerful and very fast possibilities for X-ray topographic observations, including in real time, of many phenomena involving lattice distortion in single crystals. Neutron topography is a complementary technique, slow and with worse resolution, but with unique possibilities for the investigation of heavy crystals, of magnetic domains of all kinds, and of magnetic phase coexistence.

5.9 References

[1] A. R. Lang in *International Tables for Crystallography*, Vol. C (Ed. A. J. C. Wilson), Kluwer, Dordrecht **1992,** Chapter 2.7.

[2] A. R. Lang in *Diffraction and Imaging Techniques in Materials Science* (Eds. S. Amelinckx, R. Gevers, J. Van Landuyt), North Holland, Amsterdam, **1978,** pp. 623–714.

[3] A. Guinier in *X-ray Diffraction in Crystals, Imperfect Crystals, and Amorphous Bodies*, Dover Publications, New York, **1994.**

[4] L. H. Schwartz, J. B. Cohen, *Diffraction from Materials*, 2nd ed., Springer, Berlin, **1987.**

[5] E. N. Maslen, A. G. Fox, M. A. O'Keefe in *International Tables for Crystallography*, Vol. C (Ed. A. J. C. Wilson), Kluwer, Dordrecht **1992,** Chapter 6.1.1.

[6] Z. G. Pinsker, *Dynamical Scattering of X-rays in Crystals*, Springer, Berlin **1978.**

[7] A. Authier in *International Tables for Crystallography*, Vol. B (Ed. U. Shmueli), Kluwer, Dordrecht, **1993,** Chapter 5.1.

[8] Y. Epelboin, *Mater. Sci. Eng.* **1985,** *73*, 1.

[9] A. Authier in *Diffraction and Imaging Techniques in Materials Science* (Eds. S. Amelinckx, R. Gevers, J. Van Landuyt), North Holland, Amsterdam, **1978,** pp. 715–757.

[10] U. W. Arndt in *International Tables for Crystallography*, Vol. C (Ed. A. J. C. Wilson), Kluwer, Dordrecht **1992,** Chapter 4.2.

[11] D. Raoux in *Neutron and Synchrotron Radiation for Condensed Matter Studies*, Vol. 1, HERCULES (Higher European Research Course for Users of Large Experimental Systems), Grenoble (Eds. J. Baruchel, J. L. Hodeau, M. S. Lehmann, J. R. Regnard, C. Schlenker), Les Éditions de Physique, Orsay and Springer, Berlin **1993,** Chapter 2.

[12] G. E. Bacon, *Neutron Diffraction*, 3rd ed., Clarendon Press, Oxford **1975.**

[13] V. F. Sears in *International Tables for Crystallography*, Vol. C (Ed. A. J. C. Wilson), Kluwer, Dordrecht **1992,** Chapter 4.4.4.

[14] R. Scherm, B. Fåk, in *Neutron and Synchrotron Radiation for Condensed Matter Studies*, Vol. 1, HERCULES (Higher European Research Course for Users of Large Experimental Systems), Grenoble (Eds. J. Baruchel, J. L. Hodeau, M. S. Lehmann, J. R. Regnard, C. Schlenker), Les Éditions de Physique, Orsay and Springer, Berlin **1993,** Chapter 5.

[15] M. Schlenker, J. Baruchel, *Physica* **1986,** *137B*, 309.

Part III

Acoustic Microscopy

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

1 Acoustic Microscopy

1.1 Introduction

Acoustic microscopy makes it possible to image the interaction of acoustic waves with the elastic properties of a specimen with submicrometer resolution [1]. Figure 1 shows the principle of the kind of lens that is used for high resolution work. The body of the lens is a flawless single crystal of sapphire, with its *c*-axis accurately aligned parallel to the optic axis. The top surface is flat, and on it is grown a transducer, usually made of epitaxial zinc oxide sandwiched between two gold electrodes. In the opposite face a cavity is ground with a spherical surface. There is no equivalent in acoustics to a vacuum, and so a fluid must be used to couple the waves in the lens into the specimen. Many exotic fluids have been suggested, and a small number of them have actually been used, but for practical purposes at room temperature it is hard to find one that is better than water; even for high resolution work almost the only fluid that is better is hot water! The refractive index for waves travelling from sapphire to water is very high, and two consequences of this are that even with only a single refracting surface the aberrations are very small, while at the same time the numerical aperture of the lens can be large. Almost all acoustic microscopy is done in reflection. A radio-frequency pulse is applied to the transducer, which generates an acoustic pulse in the lens. The acoustic waves are refracted at the spherical surface, are reflected by the specimen, and pass back through the lens to generate a reflected signal at the transducer. The whole interest of acoustic microscopy lies in understanding what information the reflected signal contains about the structure and elastic properties of the specimen.

Since the acoustic waves are focused on the axis of the lens, the lens must be scanned in order to build up a picture of the specimen. For this reason the technique is sometimes called scanning acoustic microscopy, partly in order to distinguish it from scanning laser acoustic microscopy [2, 3] and scanning electron acoustic microscopy [4]; but since those are both scanning techniques too, the term is not much help in distinguishing them. A simple electronic circuit for an acoustic microscope is illustrated schematically in Fig. 2. A signal of about 100 mW is generated by a synthesizer and an amplifier, and short pulses are created by a fast solid state switch. A single-pole–double-throw switch allows the transducer on the lens to be connected to that switch when the pulse occurs, and then to be switched over to the receiving preamplifier in time for the arrival of the reflected echo. In the

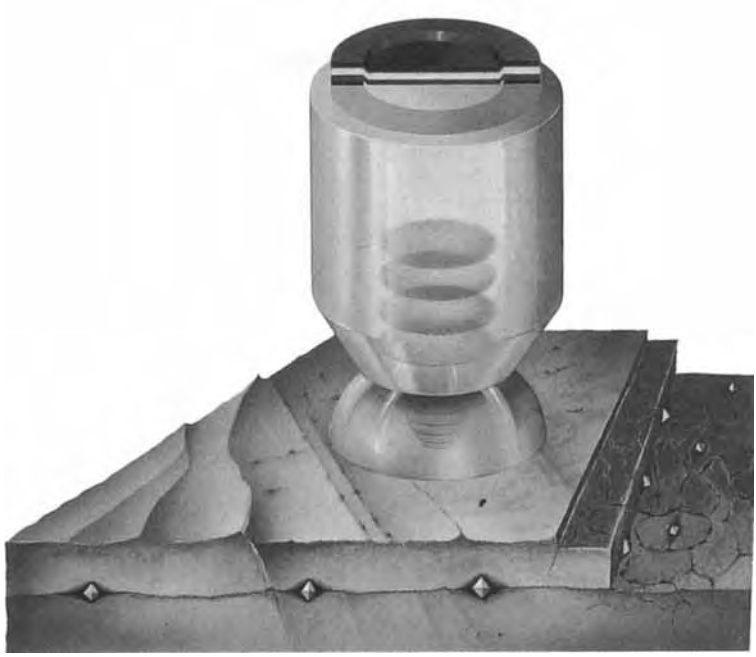


Figure 1. A lens for high resolution acoustic microscopy in reflection. The central part is a single crystal of sapphire, with the axis of the cylinder accurately parallel to its *c*-axis. The structure at the top is an epitaxially grown zinc oxide transducer sandwiched between two gold electrodes. The shaded areas within the sapphire represent the plane wavefronts of an acoustic pulse; they are refracted at the lens cavity so as to become spherical in the coupling fluid. The hemispherical feature between the lens and the specimen is a drop of the coupling fluid. The radius of the cavity would be 40 μm for use at 2 GHz (Courtesy of Leica, Wetzlar).

simple heterodyne circuit illustrated, the signal is detected using standard radio receiver technology, and its envelope is then measured by a peak detector circuit. This is the signal that would be used to modulate the brightness of each point on the picture in straightforward imaging applications. For quantitative applications the principles remain the same, but more sophisticated circuits may be used. Figure 3 shows a commercially available acoustic microscope (KSI 100). By using advanced computer control and modular circuit boards, acoustic microscopes are now available that offer exceptionally good value. This particular model has an operating frequency of 100 MHz, but a range of frequencies is available and there is

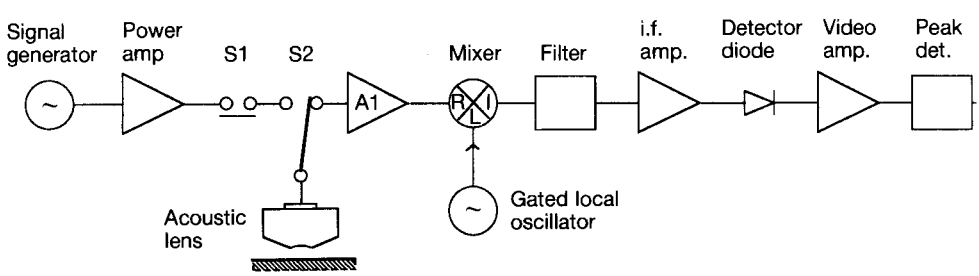


Figure 2. A simplified diagram of a basic heterodyne circuit for acoustic microscopy. The switch S1 determines the pulse length, S2 switches from transit to receive, and A1 amplifies the reflected signal. The radio frequency signal from A1 is fed to the R port of the mixer and the gated signal from the local oscillator is fed to the L port: they combine to give an intermediate frequency (i.f.) signal at the I port. The signal for either a scanned or a $V(z)$ curve would be taken from the peak detector.

Figure 3. KSI SAM 100 series acoustic microscope. This model is for use at 100 MHz and is particularly suitable for interior imaging; a wide range of frequencies is available for particular applications. All the parameters are software controlled via the computer, thus reducing cost and facilitating user control and digital data handling (courtesy of KSI, Herborn).

increasing emphasis in acoustic microscopy on choosing the acoustic frequency, and hence wavelength, that is best suited for each particular purpose.

The most obvious application of acoustic microscopy is for imaging the interior of objects—this, after all, is what ultrasonic waves are used for in nondestructive testing and in medical ultrasound. An example of this capability is shown in Fig. 4, which is a series of images of a polymer-encapsulated integrated circuit. The pictures are taken with the focal plane of the acoustic microscope at different depths below the surface. Where there is delamination between the sides of the lead frame and the plastic packaging, contrast is seen in the pictures; in frame 5 it looks bright and in frame 6 it looks dark. In this series of pictures times gating was used to select the depth that was being examined. Time-resolved measurements can be exploited in quantitative acoustic microscopy, and Sections 7 and 8 of this Chapter describe their use in the measurement of layer thickness and crack depth. Interior imaging of defects in electronic device packaging is proving to be an

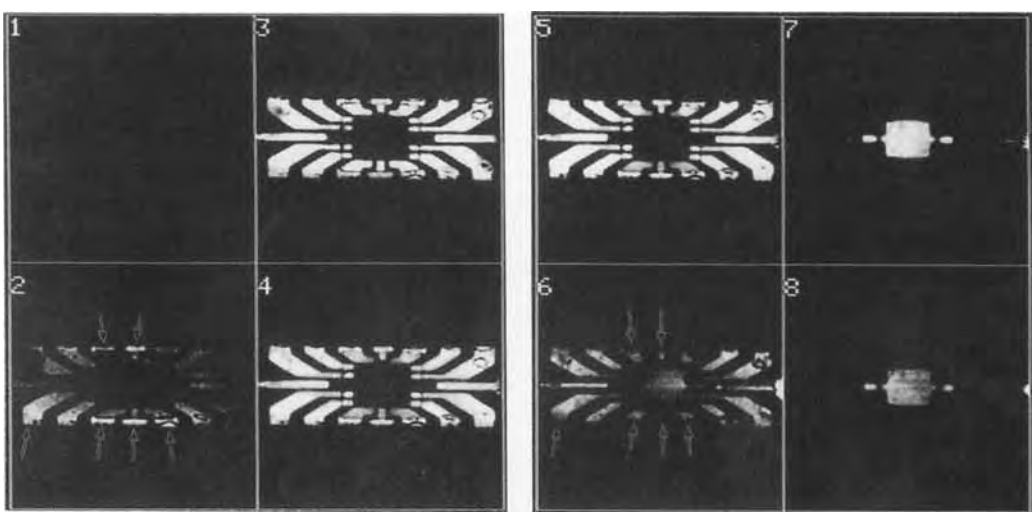
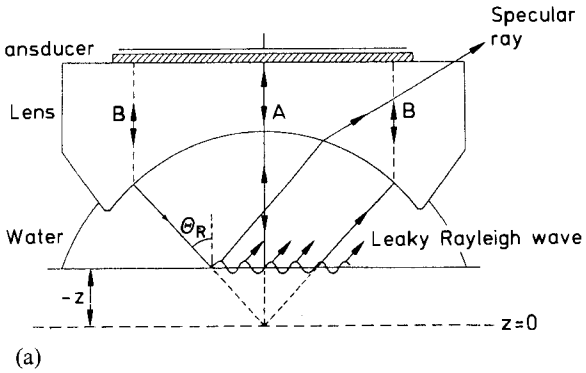


Figure 4. Acoustic images of a polymer-encapsulated integrated circuit taken at the relatively low frequency of 25 MHz, and at increasing depths. The delaminated areas at the tips of the lead frame are indicated by arrows: they are bright in frames 2–5 (where the window of the electronic time gate is above the level of the delamination), and dark in frame 6 (where the gate is below the level of delamination, so that the delamination casts a shadow in the reflected signal). The total width of each frame is 20 mm [5].

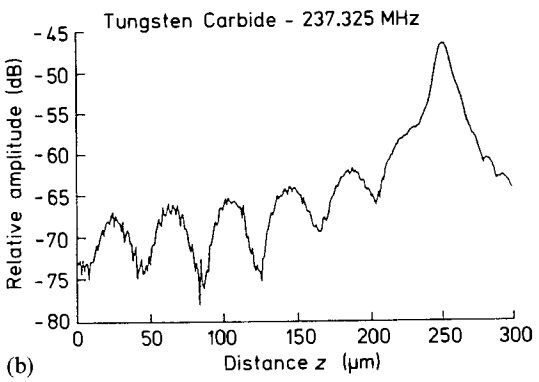
immensely important application of acoustic microscopy, and a growing range of examples is now available [6, 7]. Significant applications in biology are now also well established, especially for imaging the mechanical properties of tissue sections and of living cells [7].

There is an important effect at work which has to do with the confocal nature of an acoustic microscope. Because the acoustic waves are focused by the lens both when they travel towards the specimen and when they are reflected from it, they contribute maximum signal only when the reflecting surface is at focus. This effect is well known in scanning optical microscopy too [8], but in acoustic microscopy there can be a vital additional phenomenon. In many materials, acoustic waves incident on the specimen at an appropriate angle excite surface acoustic waves in the specimen, often known as Rayleigh waves. These waves propagate parallel to the surface with a velocity that is slower than the shear wave velocity in the bulk: they consist of a mixture of longitudinal and shear waves, each of which decays exponentially in amplitude with depth into the solid. The Rayleigh waves reradiate acoustic waves into the fluid, which return through the lens to the transducer, where interference occurs between these waves and those that have specularly reflected by simply bouncing off the surface; the phenomenon is illustrated in a ray diagram in Fig. 5a. If the lens is moved along its own axis towards the specimen, this interference causes oscillations in the detected signal. This observation is illustrated in Fig. 5b. The abscissa is the amount by which the specimen is moved relative to the focal plane of the lens, by convention this distance is denoted by z, with negative z indicating movement of the specimen towards the lens. The ordinate is the envelope detected signal: it is what would be recorded as the output of the peak detector in Fig. 2. Since it is also the video signal that would be used to modulate the intensity of an image it is denoted by V, and a plot of V as a function of z such as the one in Fig. 5b is known as a $V(z)$ curve. The oscillations which are apparent at negative z in Fig. 5b are the result of the interference between the reflection that involves Rayleigh waves in the surface of the specimen and specularly reflected waves. Figure 5c shows the amplitude and phase of acoustic waves reflected from the surface of a solid in water as a function of the angle of incidence: this is known as a reflectance, $R(\theta)$. The relationship between the reflectance function and $V(z)$ will be described in Sec. 2. The phase change of nearly 2π is associated with the excitation of Rayleigh waves in the surface of the solid: the angle at which this occurs is the Rayleigh angle, θ_R, which can be directly related to the period of the oscillations in the $V(z)$ curve, in the way to be described in Sec. 3 of this Chapter.

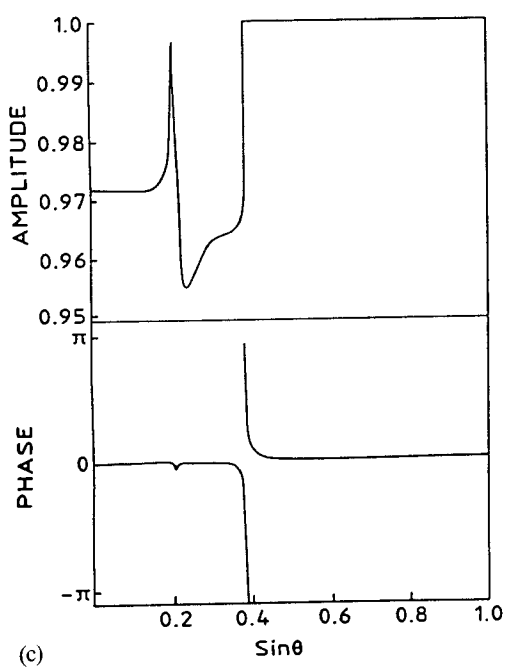
Anyone who has successfully used a microscope to image properties to which it is sensitive will sooner or later want to be able to measure those properties with the spatial resolution that the microscope affords. In an acoustic microscope such properties may include density, stiffness, anisotropy, and layered surface structure, as well as surface cracks and boundaries. In materials of high stiffness, such as most metals, semiconductors, and ceramics, a dominant role in the contrast is played by Rayleigh waves that are excited in the surface. This is emphasized particularly



(a)



(b)



(c)

dramatically when fringes of spacing half a Rayleigh wavelength are seen alongside surface cracks [9]. In specimens where Rayleigh waves play such an important role, it should be possible to measure the parameters characterizing the propagation, namely the velocity and the attenuation. In layered structures the specific propagation modes characteristic of the layers should be amenable to measurement. In lower stiffness specimens, such as polymers and biological tissue, specific surface modes may not exist, but it should nevertheless be possible to measure the propagation of acoustic waves through the material. Great advances have been made in all these areas; not surprisingly, specific techniques have been developed for specific tasks. These techniques can be used both in order to interpret quantitatively the contrast in images and as measurement techniques in their own right.

Figure 5. (a) Ray model illustrating how the signal V in an acoustic microscope at defocus can be due to the interference between two components: one due to simple specular reflection A, and the other due to a ray B, which is incident at the Rayleigh angle, excites a Rayleigh wave in the specimen, and is radiated back into the fluid as a ray symmetrical to the incident one. (b) $V(z)$ curve showing the variation of the video signal V with defocus z on tungsten carbide, at 237 MHz. The peak indicates where the specimen is at focus. The oscillations are due to the interference between reflections that involve the excitation of Rayleigh waves in the surface of the specimen and simpler specular reflections; their spacing can be directly related to the velocity of the Rayleigh waves. (c) Reflectance function, showing the amplitude and phase of the reflected acoustic waves plotted here as a function of the sine of the angle of incidence from water to a tungsten carbide surface. The phase change around $\sin\theta \approx 0.4$ is associated with the excitation of Rayleigh waves in the surface of the solid [5].

1.2 Measure of $R(\theta)$ by Inversion of $V(z)$

The variation of the video signal V is an acoustic microscope with the defocus of the lens z can be calculated from the reflectance function of a specimen, and the period and decay of oscillations in $V(z)$ can be directly related to the velocity and attenuation of Rayleigh waves. Both of these observations can be inverted in order to deduce elastic properties from measured $V(z)$ data.

If the acoustic field of an acoustic microscope is considered as made up of plane waves propagating at an angle θ to the axis of the lens, and if the pupil function of the lens and the reflectance function of the specimen are $P(\theta)$ and $R(\theta)$, and if the wave number in the fluid is k, then the response of the microscope is

$$V(z) = \int_0^{\pi/2} P(\theta)R(\theta)\, e^{i2kz\cos\theta} \sin\theta\cos\theta\, d\theta \tag{1}$$

[10–16]. The equation for $V(z)$ can be expressed as a Fourier transform by a suitable change of variables [16–19]. The new variables are defined:

$$u \equiv kz$$

and

$$t \equiv \frac{1}{\pi}\cos\theta \tag{2}$$

Then Eq. (1) may be written

$$V(u) = \int_0^{1/\pi} P(t)R(t)\, e^{i2\pi ut}\, t\, dt \tag{3}$$

With the further substitution

$$Q(t) \equiv P(t)R(t)t \tag{4}$$

$V(u)$ may be written

$$V(u) = \int_0^{1/\pi} Q(t)\, e^{i2\pi ut}\, dt \tag{5}$$

Equation (5) describes a Fourier transform [20], with $V(u)$ and $Q(t)$ as the transform pair. The limits of the integration should be from $-\infty$ to $+\infty$, but since $Q(t)$ vanishes outside the given limits this makes no difference.

The Fourier transform relationship in Eq. (5) between $V(u)$ and $Q(t)$ may be inverted to yield

$$R(t) = \frac{1}{P(t)t} \int_{-\infty}^{\infty} V(u)\, e^{i2\pi ut}\, du \tag{6}$$

Thus by measuring $V(u)$ and taking its inverse Fourier transform, the reflectance function may be deduced. Four practical constraints are immediately apparent from the theoretical formulation.

(1) Equation (6) is valid only for $1 \geq t > \cos\theta_0$, where θ_0 is the angle subtended at focus by the lens aperture; outside this range $P(t)$ in the denominator is zero. No information about the reflectance function can be obtained outside the aperture angle of the lens.

(2) The complete Fourier transform requires measurement of $V(u)$ over an infinite range. Even though $V(u)$ may be small outside the range that can be measured, the truncation will introduce errors.

(3) The inversion procedure is most straightforward when attenuation in the coupling fluid is ignored. This may

present problems in high frequency applications.

(4) The inverse Fourier transform operation must be performed using complex variables. This means that both the amplitude and the phase of $V(u)$ must be known.

Any method based on inversion of $V(z)$ actually finds the product $P(\theta)R(\theta)$, and so the first step must be to measure $V(z)$ for a material with a well-known and well-behaved reflectance function. Suitable choices are lead or PTFE (polytetrafluoroethylene, alias Teflon). A mathematically ideal reflector would have a reflectance function of unity, with no phase change, for all angles of incidence. The material that approximates most closely to this ideal is lead. Its Rayleigh velocity is too slow to allow Rayleigh waves to be excited by waves in water. Acoustic reflection is governed by the ratio of impedances of two media. In general, the greater the ratio the nearer the reflection will be to unity. Acoustic impedance may be found from the product of the density and the velocity; the common unit is the Mrayl $= 10^6$ kg $m^{-2} s^{-1}$. The longitudinal acoustic impedance of lead is 26 Mrayl (the acoustic impedance of water is approximately 1.5 Mrayl at room temperature); the modulus of the reflection coefficient is about 0.89 up to the longitudinal critical angle (43°) and almost unity above it; the phase is almost exactly zero over the whole range of practical interest. PTFE shares many of the desirable properties of lead, and it is easier to prepare and maintain a flat and clean surface on PTFE. The first critical angle for a water–PTFE interface is greater than 60°, and so the phase of its reflectance function is constant over the whole range of practical interest. The only snag is that its reflectance function is more vulnerable to variations in elastic constants. By determining $P(\theta)R(\theta)$ from the measured $V(t)$ for whatever reference material is chosen, and dividing by $R(\theta)$ calculated from the known elastic constants, the pupil function $P(\theta)$ can be determined. When $P(\theta)R(\theta)$ is subsequently determined for unknown specimens, the result can be divided by the pupil function to yield the reflectance function for that material.

The need to measure $V(u)$ as a complex-valued quantity can be met by using an accurate amplitude and phase measurement system [21]. Essentially this can consist of two synthesizers that are phase locked, with one of them acting as the master and generating the signal frequency, and the other acting as a slave and generating the local oscillator frequency and also, after division, the pulse repetition frequency. After reflection from the specimen, the signal is mixed with a gated signal from the local oscillator, and the difference is fed to a lock-in amplifier, which is referenced to the difference between the two synthesizer frequencies. If the difference is carefully chosen not to be an integer or a half-integer multiple of the pulse repetition frequency, then it is possible for the lock-in amplifier to isolate and measure the amplitude and phase of a single frequency component (within the phase noise of the synthesizers) of the signal reflected from the specimen. Because the signal frequency has passed through a switch to define the pulse, its frequency spectrum will have been broadened to the profile of the Fourier transform of the pulse shape, but then because the pulse is repetitive the profile will be broken up into a series of lines of

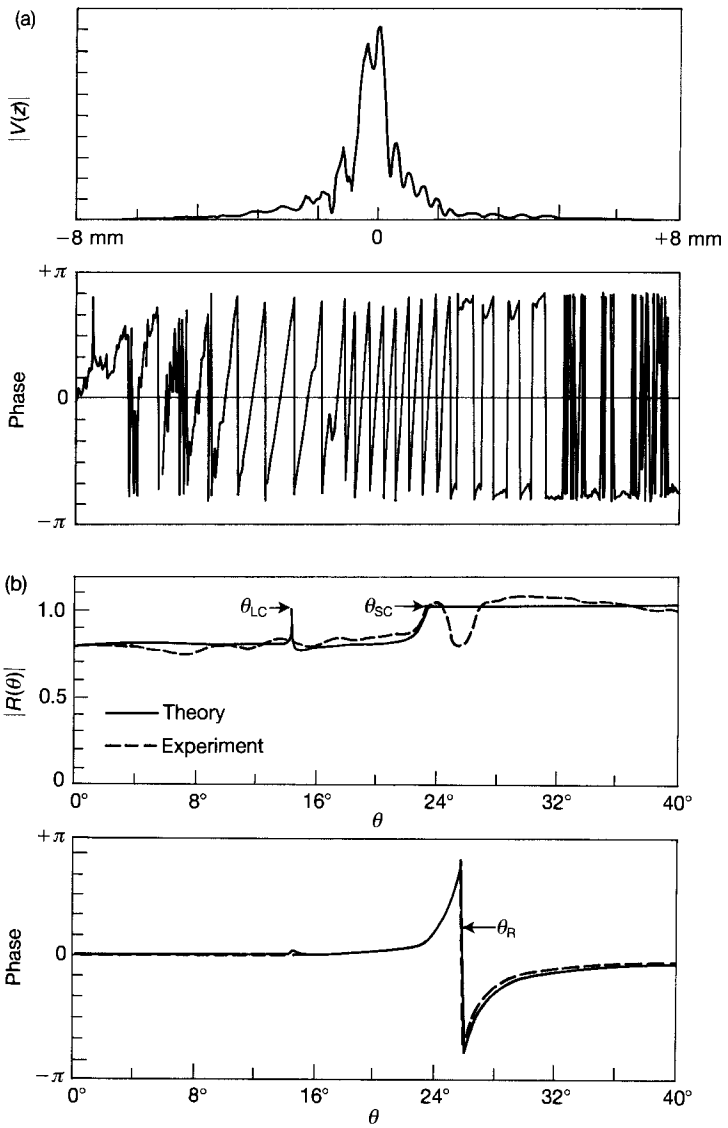


Figure 6. Fused silica: (a) magnitude and phase of an experimental $V(z)$ using a curved transducer, $\theta_0 = 45°$, frequency 10.17 MHz; (b) magnitude and phase of reflectance function – – – deduced from the experimental data in (a) using the relation $I = I_0 e^{-\mu l}$, —— calculated using the values for water, velocity $v_0 = 1486\,\mathrm{m\,s^{-1}}$; for fused silica, longitudinal velocity $v_l = 5960\,\mathrm{m\,s^{-1}}$, shear velocity $v_s = 3760\,\mathrm{m\,s^{-1}}$, density $\varrho_1 = 2200\,\mathrm{kg\,m^{-3}}$ [16].

separation equal to the pulse repetition frequency. It is one of these lines that the lock-in amplifier measures.

Figure 6 shows the modulus and phase of the reflectance function deduced from measurements made in this way using a spherical transducer at 10 MHz, together with theoretical curves calculated from the elastic constants (cf. Fig. 5b). The feature in the theoretical curves that is most precisely reproduced in the experimental results is the phase change of approximately 2π at the Rayleigh angle. This enables the Rayleigh angle, and hence the Rayleigh velocity, to be measured. The Rayleigh angle was measured to be $\theta_R = 28.85°$; using Snell's law and taking the velocity in the water $v_0 = 1486\,\mathrm{m\,s^{-1}}$

this gives the Rayleigh velocity $v_R = 3408\,m\,s^{-1}$, which compares with the theoretical value of $3415\,m\,s^{-1}$. The rise to unity in the modulus of the reflectance function also appears to be reproduced, enabling a shear critical angle $\theta_s = 23.5°$ and hence a shear velocity $v_s = 3727\,m\,s^{-1}$ to be deduced. In some cases, for example aluminum, features can be seen at the longitudinal critical angle, but in the measurements here neither the kink in the phase nor the rise to unity in the modulus is reproduced from the experimental data.

The most marked deviation from the theoretical curves occurs in the modulus at the Rayleigh angle, where there is a pronounced dip in the curve deduced from the measured $V(z)$. There are various reasons why a dip may occur in the modulus of the reflectance function at the Rayleigh angle. If the material is lossy, there may be a dip associated with attenuation of the Rayleigh wave [22]; if there is a surface layer with shear velocity slower than the substrate there may be a dip associated with the excitation of a leaky pseudo-Sezawa wave [23, 24]; if it anisotropic, there may be a dip due to phase cancellation in different directions [25]. More than one of these effects may be combined. These effects are all genuinely present in the reflectance function of the material itself. However, there is a further reason why a dip may be present in a reflectance function calculated from a Fourier inversion produce, and that is because of the limited extent of the scan in z that is available. The true reflectance function becomes convolved with the transform of the window, and this leads to phase cancellation, and therefore a dip, where the convolution occurs at a region of rapid phase change in $R(t)$. Since $V(u)$ is multiplied by a rectangular function, in the transform $R(t)$ becomes convolved with a sinc function, giving a broadening

$$\Delta\theta = \frac{4\pi}{u_1 \sin\theta} \tag{7}$$

where u_1 is the range of kz. At the Rayleigh angle in Fig. 6b, this gives a broadening $\Delta\theta \approx 2.5°$. The phase cancellation caused by the convolution process where the phase is changing rapidly causes the dip that appears in the deduced modulus. The broadening is greater the smaller the value of θ at which the phase change occurs. Thus fast materials, such as ceramics, will show even greater pathologies in the reconstruction of $R(\theta)$ around the Rayleigh angle. Oscillations associated with the sinc function can be removed by applying a smooth apodization to the $V(t)$ data, with a corresponding degradation of the resolution in the θ domain. The problems associated with the limited extent of z become more severe as the frequency f is increased, because of the f^2 increase in attenuation in water and the consequent reduction in the working distance of lenses.

In most acoustic microscopes the video signal is measured after detection by an envelope detector, so that the phase information is lost. In order to attempt to reconstruct $P(\theta)R(\theta)$ from a modulus only $V(z)$, a phase retrieval algorithm must be used [19]. The method is based on the Gerchberg–Saxton algorithm [26, 27]. Equation (5) may be written

$$V(u) = \mathcal{F}\{Q(t)\} \tag{8}$$

$\mathcal{F}$ representing the Fourier transform operator and $\mathcal{F}^{-1}$ its inverse.

The algorithm proceeds as follows. For the mth loop of the iteration, the best

estimate of $Q(t)$ is Fourier transformed,

$$V'_m(u) = \mathcal{F}\{Q_{m-1}(t)\} \tag{9}$$

This is forced to have the measured amplitude $|V(u)|$, while retaining the phase given by the algorithm

$$V_m(u) = |V(u)|\, e^{i\,phase\{V'_m(u)\}} \tag{10}$$

The inverse Fourier transform is taken to give a new estimate of $Q(t)$

$$Q'_m(t) = \mathcal{F}^{-1}\{V_m(u)\} \tag{11}$$

Finally, $Q_m(t)$ is truncated outside its allowed range by setting

$$Q_m(t)\begin{cases} = Q'_m(t), & 0 < t \leq \dfrac{1}{\pi} \\ = 0, & \text{otherwise} \end{cases} \tag{12}$$

In practice the allowed range of t may be further restricted by the pupil function of the lens. The new value of $Q(t)$ is used for the next iteration through Eqs. (9) to (12), and so on until satisfactory convergence is obtained.

The measured $V(z)$ may be improved by first tidying up the raw data. The autocorrelation function of $Q(t)$ is

$$QQ(t) \equiv \int_0^{t'} Q(\alpha)Q(\alpha + t)\, d\alpha \tag{13}$$

where the limits of integration represent the range of t outside which $Q(t)$ is zero. Hence the allowed range of $QQ(t)$ is $-t' \leq t \leq t'$. Because $Q(t)$ and $V(u)$ are a Fourier transform pair,

$$QQ(t) = \mathcal{F}\{|V(u)|^2\} \tag{14}$$

Thus the data can be preprocessed by taking the Fourier transform of $|V(u)|^2$, applying a window corresponding to the extent of the pupil function of the lens, and then taking the inverse transform to obtain a filtered $|V(u)|$, which can then be used as the data for the Gerchberg–Saxon algorithm.

Steps in the reconstruction of $P(\theta)R(\theta)$ for duraluminum alloy are illustrated in Fig. 7. $V(z)$ was measured at 320 MHz with a range of z of $\pm 240\,\mu m$ relative to

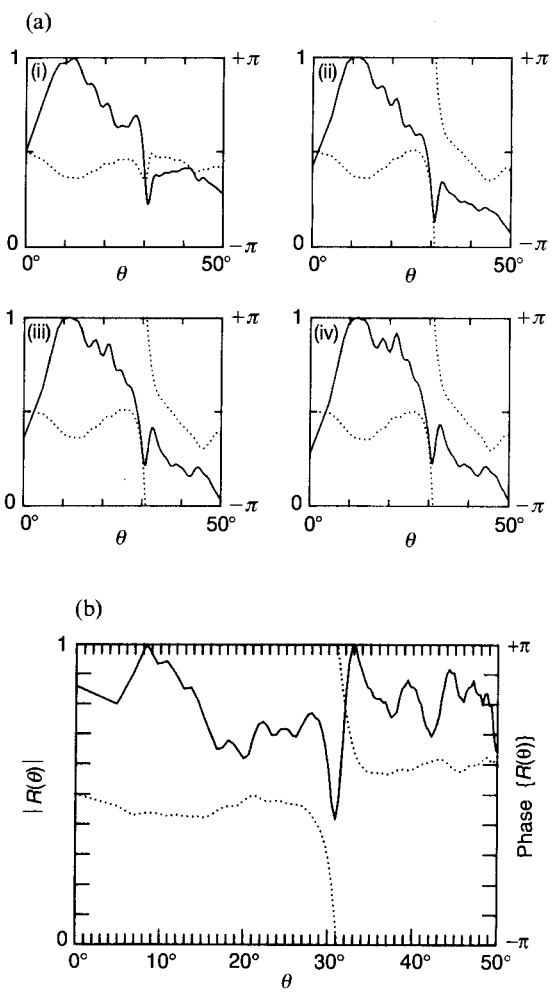


Figure 7. Reconstruction of the reflectance function of duraluminum from magnitude-only $V(z)$ data measured at 320 MHz; —— magnitude (left ordinate), · · · phase (right ordinate): (a) steps in the reconstruction of $P(\theta)R(\theta)$ after 1, 3 10, and 30 iterations of the phase retrieval algorithm; (b) reconstructed $R(\theta)$ [19].

focus. Reconstructed curves of $P(\theta)R(\theta)$ after 1, 3, 10, and 30 iterations are shown in Fig. 7a, plotted as functions of θ for the sake of familiarity. For this material, with water as the coupling fluid, the Rayleigh angle is approximately 31°, and an incipient feature is present there even after the first cycle. After 3 cycles, a phase change of 2π has developed, accompanied by a dip in the modulus at that angle. The curves after 10 and 30 iterations show little further change, and thereafter the curves remain almost constant at this level of presentation. It is remarkable to achieve such stable convergence with a one-dimensional phase reconstruction.

A pupil function $P(\theta)$ was determined for the lens using the same processing method. PTFE was used as the reference material. $V(z)$ was measured for a specimen of PTFE, and was then processed as above, to yield $P(\theta)R(\theta)$. This was divided by the calculated $R(\theta)$ for PTFE to give a calibrated pupil function. Finally, the reconstructed $P(\theta)R(\theta)$ for the duraluminum specimen was divided by this calibrated $P(\theta)$ to give a reflectance function $R(\theta)$ for duraluminum. The result is presented in Fig. 7b. The dominant feature in the result is the phase change of 2π associated with the Rayleigh angle. The angle at which this appears is in reasonable agreement with the reflectance function obtained for similar material by inversion of $V(z)$ measured with amplitude and phase information (Fig. 6). There is also a dip in the modulus at this angle; once again this is due to a finite range in z convolved with a region of rapid phase change.

If $R(\theta)$ in Fig. 1b is compared with calculated reflectance functions (cf. Fig. 5c) it is apparent that the phase change at the Rayleigh angle is the feature that is reproduced by far the most faithfully. This is because this feature corresponds to the strongest interaction of the acoustic waves in the acoustic microscope with the specimen itself.

1.3 Rayleigh Wave Measurement

Since the most prominent feature in most $V(z)$ curves, after the central maximum at focus, is the series of oscillations at negative defocus associated with Rayleigh wave excitation, it is not surprising that the most accurate information in the reconstruction of $R(\theta)$ concerns the Rayleigh velocity. The period of the oscillations in $V(z)$ is

$$\Delta z = \frac{\lambda_0}{2(1-\cos\theta_R)} \tag{15}$$

[28, 29], with the Rayleigh angle θ_R given by Snell's law

$$\sin\theta_R = \frac{v_0}{v_R} \tag{16}$$

where v_0 is the wave velocity in the fluid, and v_R is the velocity of the Rayleigh wave in the surface of the specimen. The expression for the period of the oscillations in $V(z)$ is of fundamental importance.

The change in the total attenuation suffered by the Rayleigh ray at defocus z is

$$\Delta\alpha = 2z(\alpha_0 \sec\theta_R - \alpha_R \tan\theta_R) \quad z < 0 \tag{17}$$

where α_0 is the attenuation in the fluid and α_R is the Rayleigh wave attenuation. The amplitude of the Rayleigh ray therefore varies as $\exp[-2z(\alpha_0 \sec\theta_R - \alpha_R \tan\theta_R)]$.

If the Rayleigh contribution to $V(t)$ is smaller than the geometrical contribution, the amplitude of the oscillations in $V(z)$ will follow the same exponential decay.

The expression for Δz, Eqs. (15) and (16), may be inverted to give the Rayleigh velocity in terms of Δz, the frequency f, and the velocity in the coupling fluid v_0:

$$v_R = v_0\left\{1 - \left(1 - \frac{v_0}{2f\Delta z}\right)^2\right\}^{-1/2} \tag{18}$$

Hence, by measuring the period of the Rayleigh oscillations, the Rayleigh velocity may be deduced directly. Likewise, the expression for the exponential decay of the Rayleigh oscillations can be inverted to give the normalized Rayleigh wave attenuation

$$\alpha_N \equiv \frac{\alpha_R \lambda_R}{2\pi} = \frac{(\alpha_0 \sec\theta_R - \alpha/2)\lambda_R}{2\pi \tan\theta_R} \tag{19}$$

Writing the normalized attenuation in this way enables it to be used as an imaginary component of a factor multiplying the wavenumber in an expression for wave propagation.

In the ray analysis of $V(z)$ curves, illustrated in Fig. 5a, two rays are considered to contribute to the signal: the geometrical ray and the Rayleigh ray [30]. Their contribution may be designated $\boldsymbol{V}_G(z)$ and $\boldsymbol{V}_R(z)$, using bold type to emphasize that these are complex valued quantities with phase as well as amplitude. If the measured signal is detected without phase, then its value is given by adding $\boldsymbol{V}_G(z)$ and $\boldsymbol{V}_R(z)$ using the cosine rule:

$$|V(z)|^2 = |\boldsymbol{V}_G(z)|^2 + |\boldsymbol{V}_R(z)|^2 + 2|\boldsymbol{V}_G(z)|\,|\boldsymbol{V}_R(z)|\cos\phi(z) \tag{20}$$

where $\phi(z)$ is the phase angle between $\boldsymbol{V}_G(z)$ and $\boldsymbol{V}_R(z)$. Isolating the quantity of interest, we find

$$|\boldsymbol{V}_R(z)|\cos\phi(z) = \frac{|V(z)|^2 - |\boldsymbol{V}_G(z)|^2 - |\boldsymbol{V}_R(z)|^2}{2|\boldsymbol{V}_G(z)|} \tag{21}$$

As with the methods of analysis based on Fourier inversion, it is necessary to characterize the lens response. This is again performed using a specimen in which waves are not excited in the surface. Lead is again suitable, and the measured curve is denoted $|\boldsymbol{V}_L|$. $|\boldsymbol{V}_L|$ approximates closely to $|\boldsymbol{V}_G|$. Everything on the right side (RHS) of Eq. (21) is then known, except for the function $|\boldsymbol{V}_R(z)|$. But within the ray approximation it must be a decaying exponential of the form

$$|\boldsymbol{V}_R(z)|^2 = a\exp(2\alpha z) \tag{22}$$

with only two unknowns, a and α. As a first approximation, acceptable values can be found by requiring the RHS of Eq. (22) to be an exponentially decaying sinusoidal curve that is symmetrical about the z axis. A Fourier transform of the resulting curve is taken, and numerical methods are used to find a calculated curve of the form of the left hand side (LHS) of Eq. (21) whose Fourier transform (with the same windowing) gives the best fit to the Fourier transform of the RHS [31]. This process may be repeated iteratively to converge on more accurate values for a and α, and hence enable α_N to be deduced from Eq. (19). From the period of oscillatory curve that gives the best fit, Δz and hence via Eq. (18) v_R may be deduced.

In all this processing there is scope for considerable operator skill in selecting the range of the data in each domain. Initial

filtering, for example to remove rapid oscillations of period half the water wavelength, $\lambda_0/2$, should be performed on the square of the measured data, $|V(z)|^2$, for reasons analogous to the discussion of Eqs. (13) and (14). The range of z over which data is used is determined by at least two factors. The ray approximation that the phase difference between $V_G(z)$ and $V_R(z)$ is a linear function of z, $\phi(z) = 2kz\cos\theta_R - \pi$, is not valid too near focus, and it is good practice to use data only beyond negative defocus $z \lesssim -\Delta z$, or perhaps even $z \lesssim -1.5\Delta z$. At the far negative z end, oscillations due to lateral waves may begin to dominate Rayleigh wave oscillations, and that part of the curve should be discarded for accurate Rayleigh wave measurement. On the other hand, if lateral waves are of interest, or indeed if they are the only waves present, then the different nature of the decay of lateral waves must be taken into account in the analysis.

In the form given so far, the method of analysis may be used for a standard spherical lens as used for imaging in a microscope, though because the working distance of the lens scales as $1/f^2$, a larger number of oscillations, and therefore greater accuracy, can be obtained with lower frequency lenses. But many materials are anisotropic, so that the velocity of a surface wave depends on the direction of propagation. Some useful measurements can be made on anisotropic surfaces using a standard spherical imaging lens [32]. For example, the stiffness of heavily ion-implanted silicon was found to be reduced by about 30% [33], and the elastic properties of berlinite ($AlPO_4$) crystals prepared in various acid solutions have been characterized [34]. But the most accurate measurements on anisotropic surface are made using a lens with a cylindrical surface [35], which enables measurements to be made in one direction at a time. The cylindrical lens produces a so-called line-focus beam, parallel to the axis of the cylindrical surface, and surface waves are excited in a direction perpendicular to the line focus [36]. In a line-focus-beam lens the geometrical contribution V_G is larger than the Rayleigh contribution V_R. By treating V_R as a small perturbation on V_G, the approximation may be made,

$$|V| - |V_L| = |V_R|\cos\phi \tag{23}$$

This means that a linear subtraction may be used, and there is no need for the kind of iteration that is called for by Eq. (21). Of course, that does not mean that the analysis is easy, and an immense amount of work has gone into developing an analysis algorithm that will give the highest accuracy [37]. In essence, when a series of Fourier transforms has been used to optimize the purity of the data to correspond to the RHS of Eq. (23), a final Fourier transform is performed to display the data $\mathcal{F}(\xi)$ along a reciprocal axis $\xi = 1/\Delta z$. The peak gives Δz, and hence the velocity, immediately. Since the data are taken from only a limited range of z, they can be thought of as data taken from an infinite range multiplied by a rectangle function of width ζ, where ζ is the useful range of the measurement of $V(z)$. Now if the oscillations in $V(z)$ were of constant amplitude, with no exponential decay, their Fourier transform would be of the form $(\sin\pi x)/(\pi x)$, with a peak at $\xi_0 \equiv 2\pi/\Delta z$ and zeros at $\xi_0 \pm 2\pi/\zeta$. But since there is an exponential term in the oscillations, then the zeros are partially filled in. The decay α in the Rayleigh

oscillations is related to the values at ξ_0 and at $\xi_0 \pm 2\pi/\zeta$ by

$$\alpha = \frac{2\pi}{\zeta} \frac{|\mathscr{F}(\xi_0 \pm 2\pi/\zeta)|}{[|\mathscr{F}(\xi_0)|^2 - |\mathscr{F}(\xi \pm 2\pi/\zeta)|^2]^{1/2}} \tag{24}$$

Hence the Rayleigh velocity and normalized attenuation can be calculated:

$$v_R = v_0 \left[1 - \left(1 - \frac{v_0 \xi_0}{4\pi f}\right)^2\right]^{-1/2} \tag{25}$$

$$\alpha_N = \frac{\alpha \cos\theta_R + 2\alpha_0}{2k_R \sin\theta_R} \tag{26}$$

In order to achieve the ultimate accuracy, it is necessary to take great care over the accuracy of the z-translation, the frequency, and the water temperature. The z-translation depends on the mechanical stage, and it is essential to use a good one. Adequate frequency definition cannot be obtained with envelope detection of a gated pulse, since the very act of gating introduces frequency spread. A quasi-monochromatic system must be used, either along the lines of the system described for amplitude and phase measurement in Sec. 2 of this Chapter (but now discarding the phase information) or else using a spectrum analyzer and tracking generator with zero frequency scan. The temperature measurement is perhaps the most tricky of all. There can easily be a temperature differential of a degree or two between the surface of a drop of water and its center (cf. a wet-and-dry hygrometer). At 20 °C the variation of velocity in water with temperature is $3.1\,\mathrm{m\,s^{-1}\,K^{-1}}$, giving an error of 1 part in 10^4 in velocity for a temperature uncertainty of 0.09 K. With care and patience it is possible to approach an accuracy of 0.02% in the measurement of Rayleigh velocity, and a few percent in attenuation.

It is possible to make some approximate deductions about the elastic properties of the surface being examined from the measured Rayleigh wave parameters. If the attenuation is due mainly to leaking into the coupling fluid then it may be related to the ratio of the densities of the fluid and the solid ϱ_0/ϱ_1 by [38]

$$\alpha_R \lambda_0 \approx \frac{\varrho_0 v_0^2}{\varrho_1 v_R^2} = \frac{\varrho_0}{\varrho_1} \sin^2\theta_R \tag{27}$$

The Rayleigh velocity is related to the density and the elastic constants by a sextic equation involving both the longitudinal and the shear bulk velocities. It is therefore not possible to deduce the elastic constants, even when the specimen is isotropic and the density is known, from the Rayleigh velocity alone. However, the solution to the sextic equation may be approximated by a polynomial expansion in Poisson's ratio σ [39], giving the Rayleigh velocity in terms of the density and the shear modulus μ,

$$v_R \approx \frac{1}{R_\sigma} \left(\frac{\mu}{\varrho_1}\right)^{1/2}$$

where

$$R_\sigma \equiv 1.14418 - 0.25771\sigma + 0.126617\sigma^2 \tag{28}$$

Table 1 gives values of R_σ for different values of Poisson's ratio. R_σ varies only slowly with σ, so that if the value of Poisson's ratio is known approximately, an estimate of the shear modulus can be made. In the third column of the table the percentage error in the shear modulus is given for an error of 0.01 in σ. For example, if Poisson's ratio of a ceramic material

Table 1. Values of R_σ from Eq. (29) for different values of Poisson ratio σ. The third and fourth columns give the percentage error in the shear modulus μ and Young modulus E for an error in σ of 0.01

σ	$1/R_\sigma$	$\Delta\mu/\mu$ (%)	$\Delta E/E$ (%)
0.0	0.8740	−0.45	0.55
0.1	0.8931	−0.42	0.49
0.2	0.9110	−0.38	0.46
0.3	0.9274	−0.34	0.43
0.4	0.9422	−0.29	0.42
0.5	0.9551	−0.25	0.42

were expected to be 0.2 ± 0.01 (an uncertainty of 5%), the shear modulus could be deduced from the Rayleigh velocity to an accuracy better than 0.4% (plus any error in the measurement and in the value of the density). Young's modulus of an isotropic material is related to the shear modulus and Poisson's ratio by

$$E = 2\mu(1 - \sigma) \tag{29}$$

The error in Young's modulus for an error of 0.01 in σ is given in the fourth column of the table. In this case an estimate of $\sigma = 0.2 \pm 0.01$ would give an uncertainty of less than 0.5% in E.

1.4 Anisotropy

Because shear waves in anisotropic materials are not degenerate, both Rayleigh and pseudo-Rayleigh (or pseudo-surface) waves can exist in some anisotropic surfaces. Like Rayleigh waves, pseudo-Rayleigh waves have both longitudinal and transverse components, and the transverse components of Rayleigh and pseudo-Rayleigh waves are orthogonal and nondegenerate. Unlike the Rayleigh wave, the pseudo-Rayleigh wave is not completely bound to the surface even in the absence of a fluid; it is slightly faster than the bulk shear wave of the other polarization, and it is able to couple weakly into that shear wave and so lose energy into the bulk. Rayleigh and pseudo-Rayleigh waves can be observed and measured as a function of angle using a cylindrical lens with the line-focus-beam technique. Figure 8a is a measured $V(z)$ curve for a GaAs (001) surface, with propagation direction in the surface $\phi = 26°$ relative to [100]. At this angle both

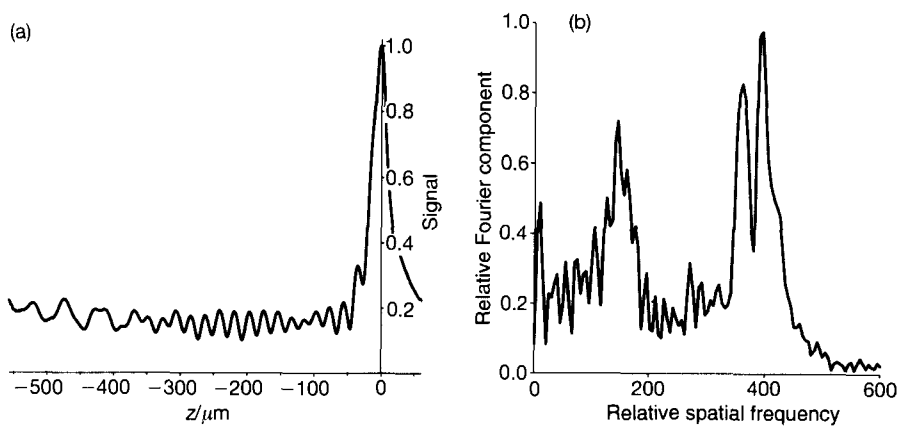


Figure 8. GaAs (001), $\phi = 26°$: (a) $V(z)$, measured at 225 MHz; (b) Fourier transform of the $V(z)$ data (courtesy of Sue Spencer).

Rayleigh and pseudo-Rayleigh waves are strongly excited. Beating between the two modes is apparent, with cancellation occurring around $z \approx -110\,\mu m$, and partial cancellation beyond $-300\,\mu m$ (the two waves do not suffer the same rate of decay). At sufficiently large defocus, oscillations due to longitudinal head waves also appear. The individual contributions can be analyzed in the Fourier transform of $V(z)$, Fig. 8b. The slowest mode (i.e., the peak at the lowest values of k) is the longitudinal head wave; the peaks at higher k are the pseudo-Rayleigh and Rayleigh waves, close together but nevertheless adequately separated.

In order to measure the velocity and attenuation of each wave using the method of Sec. 3, the peaks must be analyzed one at a time. The results of a set of such measurements on a GaAs (001) surface, for ϕ from 0° to 90°, are presented in Fig. 9. For $\phi < 20°$ only the Rayleigh wave could be measured. In the range $20° \leq \phi \leq 35°$ both Rayleigh and pseudo-Rayleigh waves were detected. Above 35° only the pseudo-Rayleigh wave is detected. The solid curves in Fig. 9 are the theoretical curves for the Rayleigh and pseudo-Rayleigh velocities in the absence of fluid loading, together with the curve for the quasi-shear horizontal-polarized bulk wave which they approach asymptotically. GaAs (110) exhibits no pseudo-Rayleigh waves, but the coupling to Rayleigh waves is stronger and there is only one mode present in the analysis. On a GaAs (111) surface, both Rayleigh and pseudo-Rayleigh waves are present at all angles but the pseudo-Rayleigh wave does not couple to the fluid in the ⟨110⟩-type symmetry directions.

As well as semiconductor wafers, studies have been made of piezoelectric wafers by the line-focus-beam technique, in particular lithium niobate ($LiNbO_3$). Lithium niobate is an important optoelectronic material, because light beams in it can be switched by exciting surface

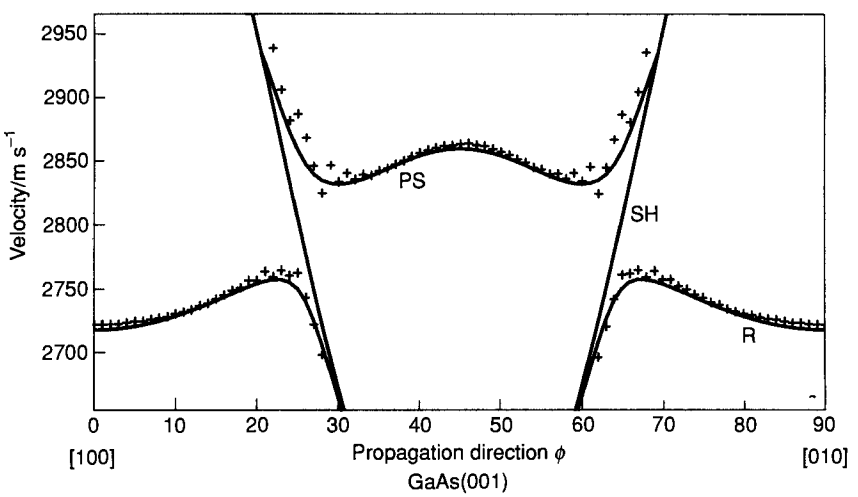


Figure 9. Rayleigh and pseudo surface wave velocities measured on GaAs (001), indicated by +; the propagation direction ϕ is measured relative to a ⟨100⟩ direction. The solid lines are calculated curves of the Rayleigh and pseudo-Rayleigh velocities (without fluid loading), and the quasi-shear horizontal polarized bulk wave which they approach asymptotically (Courtesy of Zenon Sklar).

acoustic waves in the surface so that the light suffers Bragg scattering. One way of guiding the light is to remove lithium atoms from the material in the vicinity of the surface, exchanging them for protons to maintain charge neutrality. The proton-exchanged material has a slower velocity of light (i.e., a higher optical refractive index), and so light can be confined to a channel, for much the same reason that a Rayleigh wave is confined to a surface, or indeed that any conventional waveguide works. But if the proton-exchanged material is to be used with surface acoustic wave switching, it is essential to know how the elastic properties are affected.

Measurements were made of the three principal crystallographic surfaces, and the effects were quite large [40]. The behavior was most straightforward in the Z-cut face, where the proton exchange caused a reduction in Rayleigh velocity of about $200\,\mathrm{m\,s^{-1}}$ in all propagation directions. The effects in the X-cut face were more complicated, but the maximum velocity change was about the same. In the technologically important Y-cut face, proton exchange had little effect on Rayleigh wave propagation within 45° of the Z-propagation direction. Beyond that there was a decrease in Rayleigh velocity relative to the virgin material: in the X-propagation direction the difference was $140\,\mathrm{m\,s^{-1}}$. The attenuation was also measured. The decrease in velocity within 45° of the X-propagation direction was accompanied by an increase in Rayleigh attenuation: in the X-propagation it was double its value in virgin $LiNbO_3$. Doping with MgO increases the optical power that $LiNbO_3$ can handle without damage, as well as increasing the refractive index, and this may make it a candidate for optical fibers. The correlation between the effects on the optical and acoustic velocities has been measured: 5% doping causes an increase in the optical refractive index of 0.7% and an increase in the Rayleigh wave velocity of 0.9% [41]. Lithium niobate is also used for surface acoustic wave filters, both for military use and in domestic television sets. Measurements have been made area by area on whole wafers, and correlated with the performance of devices previously fabricated on the wafer and removed before the $V(z)$ measurements were made [42, 43]. The variation in Rayleigh velocity was about 0.6%, and the correlation between regions on the wafer of high and low velocity exhibited by the devices and measured by $V(z)$ was excellent. Lithium tantalate and zinc oxide on glass have also been measured in this way [44, 45].

Not all anisotropic materials are perfect single crystals, and not all of them give such tidy results as electronic and optoelectronic wafers. But angle-resolved Rayleigh wave measurements can be used to characterize the anisotropy of heavily drawn metal–metal composites. Dual phase alloys of Cu–20% Nb were originally developed to investigate their superconducting properties. It turned out that they also had exceptional mechanical strength, together with good high temperature electrical conductivity. Cu–Nb composites are prepared by casting them so as to produce an array of Nb dendrites in a copper matrix. They are then extruded or rolled to break up the dendrites and align them into long ribbons of filaments or platelets with separation less than a micrometer. The local crystallographic orientations of the filaments and the matrix determine the directions in which dislocation motion can occur, and hence control

the mechanical properties. The extent of the strengthening depends strongly on the draw ratio η, which is the logarithm to base e of the initial to the final cross-sectional area.

In a series of experiments on materials of this kind in a line-focus-beam instrument, specimens of Cu–20% Nb, and a related Cu–20% Ta alloy, of varying η were measured, together with homogeneous samples of the parent metals Cu and Nb [46]. The measurements proved extraordinarily difficult, and only the drawn samples of the composite alloys gave $V(z)$ curves that could be satisfactorily analyzed to yield a Rayleigh velocity that could be trusted. Measurements were made at 5° intervals, and the results for the samples with $\eta = 3.6$ and $\eta = 5.4$ are given in Fig. 10. Theory suggests that the angular dependence of the Rayleigh velocity in single-phase polycrystalline metals with weak anisotropy should be a polynomial of the form $a + b\cos 2\phi + c\cos 4\phi$ [47]. Best fit curves of this form, with the zero of ϕ also chosen for best fit, have been drawn in Fig. 10. Of course, the specimens are not single phase, and they may be more than weakly anisotropic, nevertheless something can be made of the curves. They have been drawn with the same vertical axis, and the anisotropy in the Rayleigh velocity of the less drawn specimen is only a quarter of that of the other one. The Cu–20% Nb alloy gave considerably greater variation of v_R with ϕ than the Cu–20% Ta alloy. From the fitted curves, values of the parameters δW_{400} and W_{420} commonly used to describe texture in pole figures of X-ray or neutron diffraction measurements were calculated, and these made sense in terms of what would have been expected. It would be difficult to measure such a small specimen by either of those techniques, and it would be tedious to measure enough grain to get a sensible average by electron microscopy.

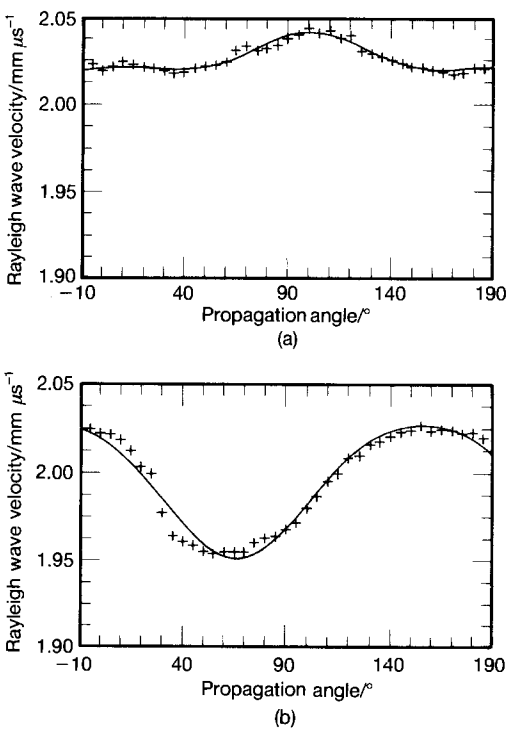


Figure 10. Angular dependence of the Rayleigh velocity in Cu–20% Nb with draw ratio (a) $\eta = 3.6$, (b) $\eta = 5.4$ [46].

For many applications it is desirable to be able to measure with both angular and spatial resolution. With a single lens function there has to be a Fourier compromise between these two. If you have the good spatial resolution of a spherical lens, then there is no angular resolution of the Rayleigh waves; they are generated and detected in all surface directions. On the other hand, a line-focus-beam lens gives very sharp definition of the azimuthal angle, but poor resolution along the length of the line. It would be possible to achieve high spatial resolution and high angular

resolution by exploiting the fact that a reflection confocal microscope is effectively a folded transmission confocal microscope, so that there could be two different pupil functions, one a full aperture to give spatial resolution, and the other a non-axially symmetric aperture to give angular resolution. This could be achieved by depositing two transducer patterns on top of one another, one for sending and the other for receiving, with a common ground electrode between them. But experiments with so-called bow-tie transducers have shown how difficult it is to obtain the desired result, chiefly because of diffraction in the body of the lens [48]. More straightforward directional point-focus lenses with the same pupil function for both transmitting and receiving have proved more successful. A lens with a strip transducer has been demonstrated that gives good azimuthal resolution at the same time as spatial resolution better than a wavelength at 225 MHz [49, 50], and other designs can be developed [51]. There may be further development and applications of such lenses for quantitative acoustic microscopy with spatial resolution.

Rather accurate measurements of stress on a microscopic scale have been made using a lens with a transducer that generates both longitudinal and shear waves. The longitudinal wave is refracted at the lens surface and generates a focused wave in the coupling fluid in the usual way. The shear wave travels more slowly. At the lens surface it generates waves in the fluid only where it excites motion normal to the lens–fluid interface. This means that it will not excite any axial ray, and also it will not excite waves where the polarization is tangential to the lens surface. The transmitted energy is therefore concentrated in two lobes on opposite sides of the axis, with an azimuthal distribution approximately proportional to $\cos^2\phi$, where ϕ is the azimuthal angle measured parallel to the shear wave polarization [52]. This field distribution offers a good compromise between azimuthal resolution and spatial resolution. The longitudinal wave excited by the transducer gives a signal that is separated in time from the shear wave signal (because the longitudinal wave travels faster in the lens), and gives an accurate phase reference because it travels through almost the same path in the fluid. There is therefore compensation for fluctuations in temperature and lens height.

Measurements of stress in a 15 μm thick layer of alumina sputtered on a glass substrate are shown in Fig. 11 [53]. From the subsequent curvature of the glass, it was estimated that there was a compressive residual stress of −40 MPa in the alumina. A trench was then etched in the alumina; the end of the trench can be seen in Figs. 11a and b. The dark broad ∩-shape 50 μm wide is the sloping wall of the trench, and the bottom of the trench, which is at the same depth as the top of the glass substrate, looks more or less the same as the flat surface of the alumina. Just beyond the end of the trench, measurements were made of the surface stresses in the x-direction, σ_{xx}, and in the y-direction, σ_{yy}. In Fig. 11a the sum of these stresses near the trench have been represented in colour and superimposed onto a photograph of the specimen. Away from the trench the sum of the stresses is approximately twice the residual stress, but close to the trench the stress is reduced to a value closer to the residual stress itself.

More quantitative comparison is possible with theoretical results from a finite

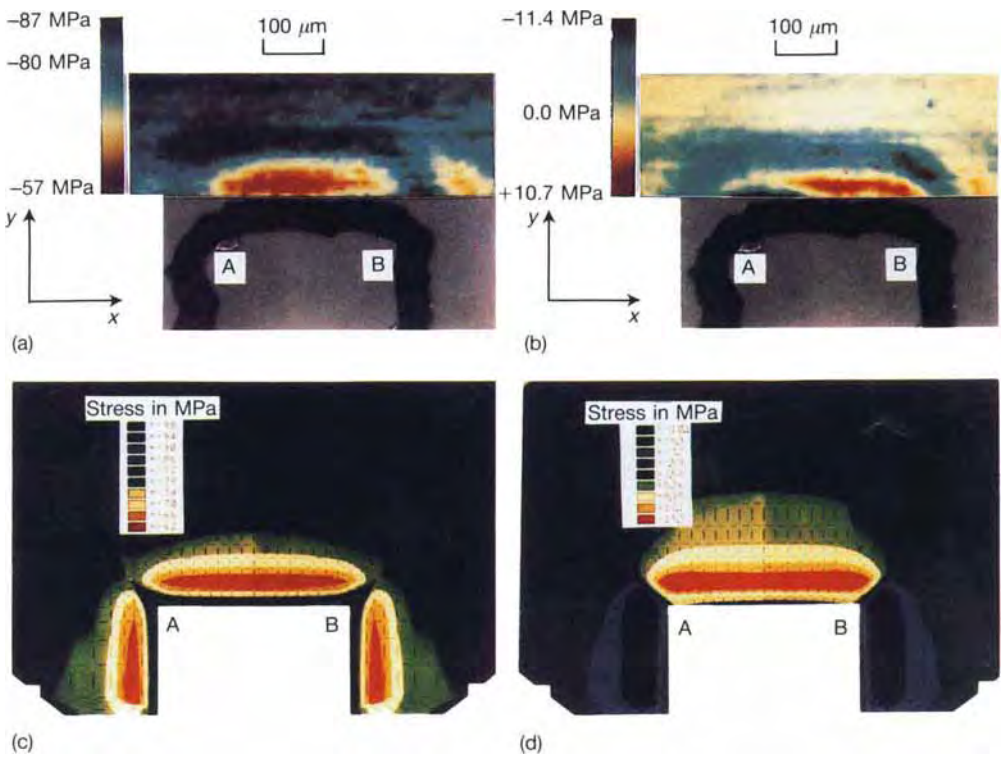


Figure 11. Stress fields at the end of a trench etched in a 15 μm thick layer of sputtered alumina on a glass substrate. The trench was 15 μm deep, 0.4 mm wide, and 10 mm long. The long range residual stress in the alumina layer measured from the curvature of the glass substrate was −40 MPa (compressive). The top two collages are photographs of one end of the trench, the dark meandering curve about 50 μm wide being the sloping wall of the trench, and the points A and B lying on the floor of the trench and marking its corners, with measurements by acoustic microscopy of (a) the sum of the stresses $\sigma_{xx} + \sigma_{yy}$ and (b) the difference of the stresses $\sigma_{yy} - \sigma_{xx}$; $f = 670$ MHz. The bottom two pictures are finite element calculations of the same geometries (neglecting the slope of the trench walls), with the points A and B corresponding to those in the upper pictures and the color scales corresponding in each case to the picture above, of (c) the sum of the stresses $\sigma_{xx} + \sigma_{yy}$ and (d) the difference of the stresses $\sigma_{yy} - \sigma_{xx}$ [53].

element calculation of $\sigma_{xx} + \sigma_{yy}$ in Fig. 11c. The color scales approximately correspond in the two pictures, and the points marked A and B in the two pictures roughly correspond to the same points on the specimen. The calibration of the velocity shift with stress was obtained by measuring the velocity difference in the two directions very close to the wall of the trench, and assuming that there would be a uniaxial stress of −40 MPa. The reduction in the sum of the stresses in the vicinity of the wall of trench might be supposed to be due to the relaxation of σ_{yy}, and this is confirmed by the map of the difference between the stresses, $\sigma_{yy} - \sigma_{xx}$, in Fig. 11b. Far from the trench, the difference is small, but close to the wall the difference becomes positive, because both components are negative (i.e., compressive), but σ_{yy} is smaller. Comparison can be made with the finite element calculation of $\sigma_{yy} - \sigma_{xx}$ in Fig. 11d. These results are a remarkable achievement.

They were obtained at 670 MHz, with a spatial resolution of 5–10 μm. The total velocity changes involved were of the order of 0.5%, and the stress sensitivity achieved was 0.3 MPa.

1.5 Surface Layers

Surface layers introduce a whole new range of surface wave modes that can be excited in an acoustic microscope, and these can be exploited in quantitative measurements to give information about the thickness and composition of the layer, and about the presence of any disbonding [54]. Measurements have been made at frequencies up to 1 GHz [55]. The simplest case is a single surface layer on a substrate. The dispersive reflectance function for a layer of gold on 42-alloy is shown in Fig. 12 [24]. The reflection coefficient (a) has been plotted upside down in order to display high coupling into the substrate as peaks in $1 - |R(\theta)|$. At zero normalized thickness $R(\theta)$ has the familiar form of a fluid-loaded half-space. As the thickness is increased so the pseudo-Sezawa wave can be seen to develop, with decreasing velocity (given by Snell's law as $v_0/\sin\theta$), until the critical angle for shear waves in the substrate, when it becomes a true Sezawa wave and ceases to couple any energy into the substrate, manifesting itself only in the phase of $R(\theta)$. Higher modes can also be seen beginning to develop at greater layer thicknesses. Just as in a uniform half-space the interaction between waves in the fluid and the solid does not stop at the shear critical angle, but is manifested in a phase change of 2π associated with Rayleigh waves, so there is also a phase change associated with Sezawa waves. In the plot of the phase of the dispersive reflectance function (b), the 2π phase change is well pronounced, especially in the region where the reflectance function is unity. The analogy with Rayleigh waves is a close one; the 2π phase change is associated with the excitation of Sezawa waves which cannot radiate energy into the solid but can couple weakly into waves in the fluid.

Pseudo-Sezawa waves have been observed in $V(z)$ data in acoustic microscopy [23], and they have also been identified in a reconstructed $R(\theta)$ by the inversion of $V(z)$ [16]. Very accurate measurements of a single layer on a substrate have been made using the line-focus-beam technique [56]. Figure 13 shows $V(z)$ curves at 225 MHz for gold films of 100 nm and 300 nm on fused silica, together with theoretical dispersion curves for the velocity and attenuation of pseudo-Sezawa waves and Rayleigh waves. In the thinner coating the Rayleigh wave is well excited, and its velocity can be measured by analyzing $V(z)$ in the range $-150\,\mu\text{m} \leq z \leq -25\,\mu\text{m}$. The pseudo-Sezawa wave is also excited, but not so strongly. In the thicker coating the pseudo-Sezawa wave is the strongest mode and its velocity can be measured by analysis of almost the whole range of z beyond $z < -75\,\mu\text{m}$. If the film is regarded as an elastic continuum with perfect surfaces, then, from the dispersion curves in Fig. 13c, a measurement sensitivity of about $1\,\text{m}\,\text{s}^{-1}$ would correspond to a sensitivity to change in film thickness of 0.4 nm for the 100 nm film and 0.2 nm for the 300 nm film. Absolute interpretation should be carefully checked against films of comparable thickness prepared in a similar way, because it can never be

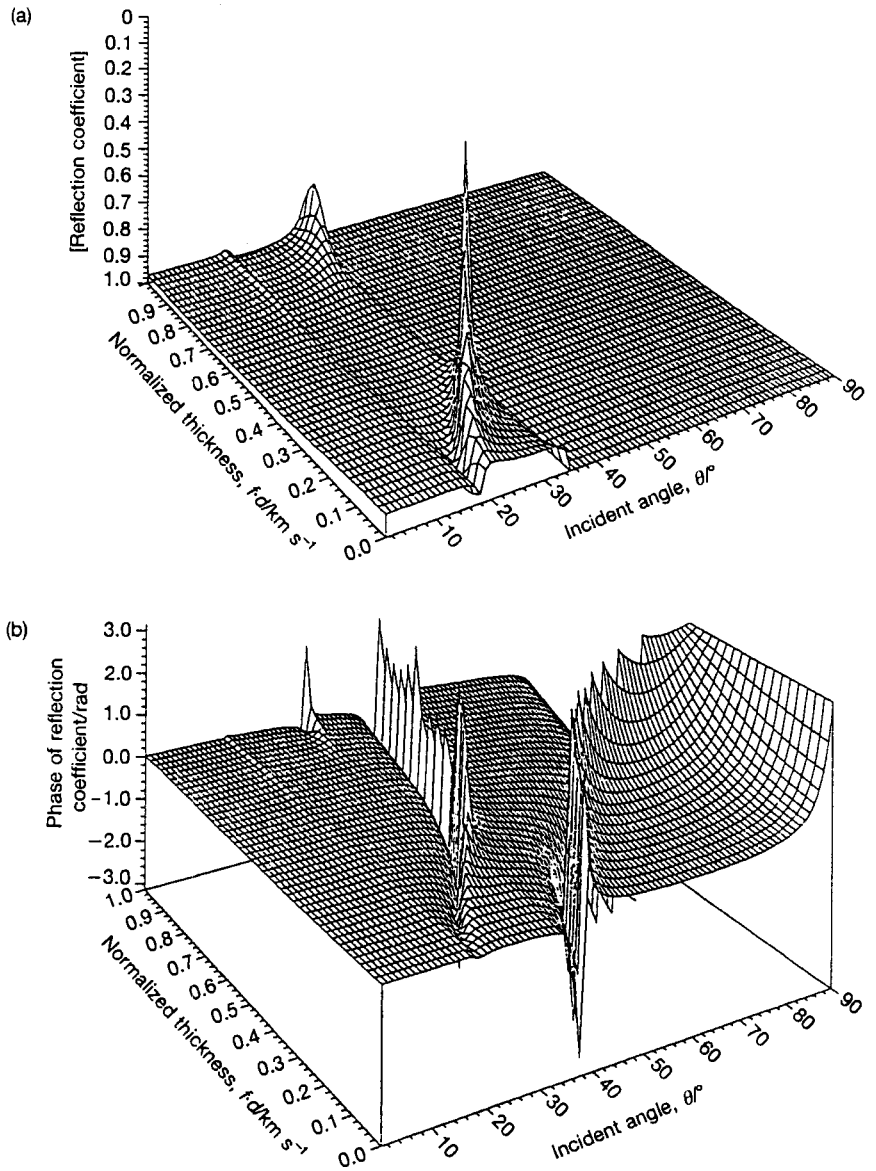
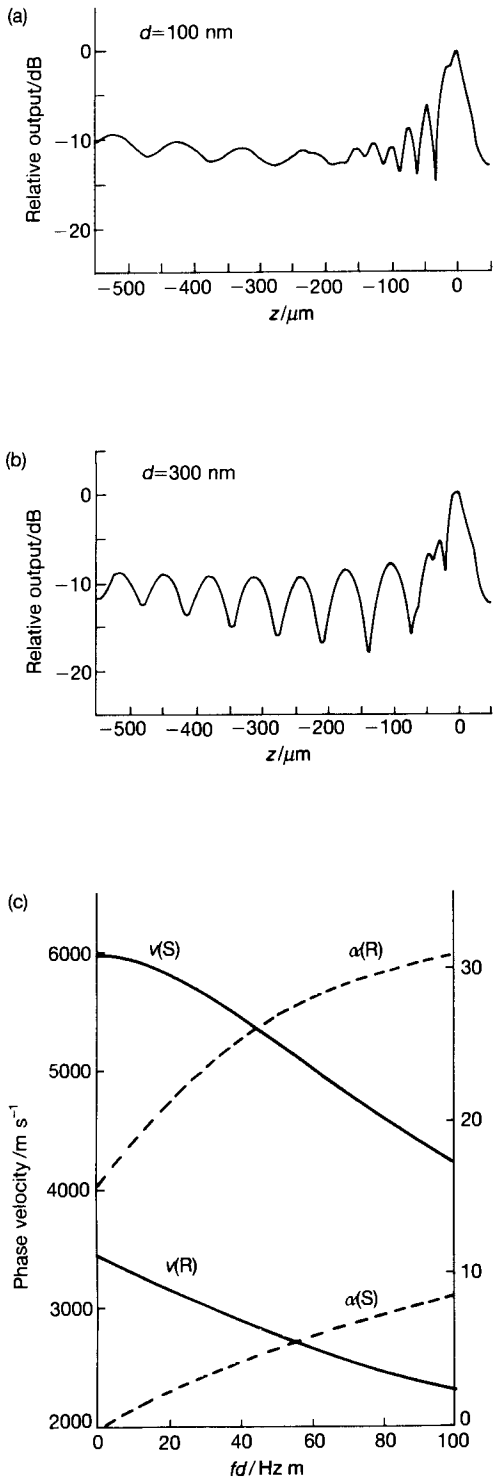


Figure 12. Reflectance function for a layer of gold on 42% Ni–Fe alloy as a function of both the incident angle θ and the frequency – thickness product fd: (a) magnitude, plotted as $1 - |R(\theta)|$ to show the sharp minima in the magnitude; (b) phase [24].

assumed that the properties of thin films are the same as those of bulk material.

There is a particularly abrupt change in the attenuation measured in an acoustic microscope at approximately the value of fd that corresponds to the transition from pseudo-Sezawa to Sezawa wave in the absence of a fluid, i.e., as the mode velocity crosses from above the shear velocity in the substrate to below it [57]. In simple



terms, below that value of fd the attenuation is due to a combination of leaking into longitudinal waves in the fluid and shear waves in the substrate, whereas above that value of fd only leaking into the fluid occurs. If the fd product for the given combination of materials is known, measurement of the frequency at which this abrupt drop in attenuation happens gives a particularly sensitive way of determining the thickness of the film.

By measuring surface wave velocity over a range of frequencies, and especially over a range of angles on anisotropic substrates, it is possible to use inversion techniques that have been established to deduce elastic properties or thicknesses of surface layers, including layers with stresses [58, 59].

It is apparent from Fig. 12 that the critical phenomena in the reflectance function can be intersected either by scanning the range of the angle or by scanning the range of the frequency. In a $V(z)$ measurement, when the data is Fourier transformed it is displayed as a function of $1/\Delta z$, the period of the oscillations in $V(z)$. With a change of variable, this is equivalent to displaying the signal as a function of angle, since θ_R and Δz are related directly by Eq. (15). Hence angles at which strong features occur can be related to velocities via Snell's law, Eq. (16). Analyzing the data effectively as a function of angle in this way can be considered as equivalent to performing a scan

Figure 13. Gold films on a fused quartz substrate: 225 MHz. (a), (b) $V(z)$ measurements: (a) 100 nm thick gold film, (b) 300 nm thick gold film. (c) Theoretical dispersion curves for leaky Rayleigh waves (R) and leaky pseudo-Sezawa waves (S): —— velocity, – – – attenuation [56].

along a line of constant fd in Fig. 13. An alternative way to measure a dispersive specimen is to use a fixed angle of incidence and sweep the frequency. At a given angle there will be phenomena such as a sharp dip in the reflectance function at the frequency at which a Sezawa wave velocity crosses the substrate shear wave velocity. This can be done using pairs of transducers in the frequency range 20–140 MHz [60]. At these relatively low frequencies it is easy to see that a reflection system has two pupil functions that can be quite independent, so that the waves from one of the transducers can be focused to give spatial resolution, while the other one remains unfocused to retain the angular definition of the planar geometry. In this way it is possible to achieve spatial resolution of 10 μm at frequencies up to 300 MHz, while retaining a 130 MHz bandwidth.

For even higher spatial resolution, including use for imaging as well as quantitative examination, special conical lenses have been developed, to give the same angle of incidence from all parts of the lens surface, with axial symmetry [61]. These lenses give exceptional sensitivity on layered specimens to such properties as coating thickness and composition, and also to disbonded areas. In one experiment, copper was plated onto steel, first with copper cyanide as the electrolyte on selected areas, and then with copper sulfate as the electrolyte over the whole specimen. The regions that were masked during the copper cyanide stage were expected to give poorer bonds, and indeed the difference between the two kinds of bond quality showed up as a shift in the frequency at which a maximum signal was obtained, from 10 MHz over the good bond to 8 MHz over the poorer bond [62]. It may be that for such examinations measurements of $V(f)$ will become as important as $V(z)$.

1.6 Surface Brillouin Spectroscopy

The highest frequency routinely used in acoustic microscopy is 2 GHz, and for the kind of quantitative surface wave measurements described in Sections 3–5, 225 MHz has been established as the most usual frequency. This corresponds to a Rayleigh wavelength of 10–20 μm, which means that a comparable depth is sampled (Table 6.2 of Ref. [1]). In many applications, such as the characterization of surface coatings or subsurface damage, the significant depth is considerably less than this, and greater sensitivity to surface elastic properties could be obtained by using a higher frequency. If the attenuation in the coupling fluid prevents the frequency being increased in conventional focused acoustic microscopy, is there another way?

In Brillouin spectroscopy, light interacts with the elastic properties of a solid to produce a frequency shift in the light that is characteristic of an elastic wave in the solid that scatters the light [63]. In quantum terms, a photon of light either creates or absorbs a phonon (Stokes and anti-Stokes respectively). In this process energy and momentum are both conserved. Conservation of energy gives a scalar equation

$$\pm\Omega = \omega_2 - \omega_1 \tag{30}$$

where Ω is the phonon frequency, ω_1 and ω_2 are the incident and scattered photon

frequencies, and + or − refer to anti-Stokes or Stokes. Conservation of momentum gives a vector equation

$$\pm \boldsymbol{K} = \boldsymbol{k}_2 - \boldsymbol{k}_1 \tag{31}$$

where $\boldsymbol{K}$ is the phonon wavevector and $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$ are the incident and scattered photon wavevectors. In surface Brillouin spectroscopy the component of wavevector parallel to the surface is conserved (cf. Snell's law), giving a scalar equation

$$\pm K = k_2 \sin\theta_2 - k_1 \sin\theta_1 \tag{32}$$

where θ_1 and θ_2 are the angles of the incident and scattered light beams. The frequency of the phonon is usually about five orders of magnitude less than the frequency of the photons, and so to a good approximation $k_2 \approx k_1$. In many experiments the scattered light is detected back in the direction of the incident light, so that $\theta_2 \approx -\theta_1$. To a good approximation we may then write

$$\pm K = 2k_1 \sin\theta_1 \tag{33}$$

Thus if we know the wavelength of the incident light and the velocity of light, and the angle of incidence, then we can deduce the wavenumber of the surface phonon. If we can measure the frequency shift of the scattered light then we immediately know the phonon frequency, and hence we can deduce the velocity of surface phonons. This is the principle of surface Brillouin spectroscopy. The same result can be derived classically by treating the elastic wave as a traveling diffraction grating which imparts a Doppler shift to the light.

It is immediately apparent from Eq. (33) that the phonon wavelength will be comparable with the wavelength of the light. The wavevector lies on an acoustic branch and is close to the center of the Brillouin zone, so that the phonon acts as a nondispersive elastic wave. For comparison, in Raman spectroscopy phonons are excited on optical branches, also near the center of the Brillouin zone, so that the frequency is almost independent of the wavevector. Thus Brillouin spectroscopy measures quasi continuum elastic properties, whereas Raman spectroscopy measures nearest neighbor local vibrations. There are several mechanisms whereby the light energy can couple to elastic waves. The strongest, when it is present, is through elasto-optic coupling in the material. But all materials can couple to light through the surface ripples which occur when a surface acoustic wave propagates, and so this mechanism is always present in surface Brillouin spectroscopy. The strength of the coupling depends on the angle of incidence. It generally falls off beyond $\theta \geq 70°$, and so an angle of incidence of 70° is often used to maximize the value of K from Eq. (33) without compromizing the strength of the Brillouin signal.

A commerically available surface Brillouin spectrometer is shown in Fig. 14 (Bede Scientific Instruments Ltd, Durham). The sample is mounted on a stage which permits scanning, rotation, and tilt in six axes, for samples as large as 150 mm semiconductor wafers. The complete optical system is mounted on a vertical plate, which can be rotated through 90° in its plane. Light of wavelength 532 nm is generated by a frequency-doubled Nd–YAG laser and is incident on the sample at an angle determined by the orientation of the plate. The scattered light is collected by a lens which is coaxial with the incident beam, and is then reflected to pass through a spatial filter and a Fabry–Perot interferometer. Brillouin scattering

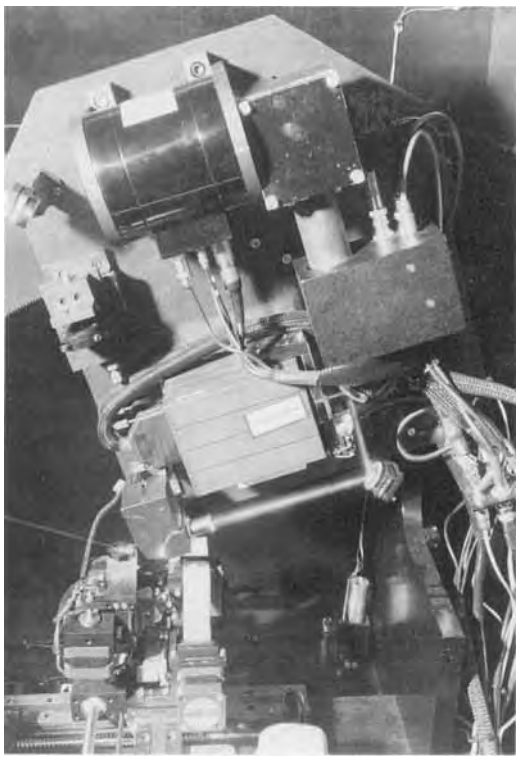


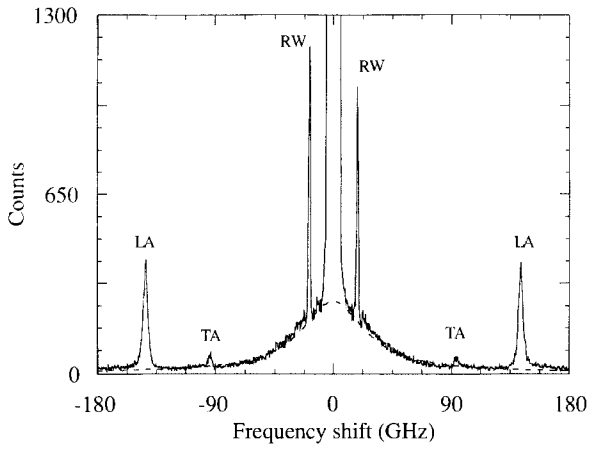
Figure 14. Bede BriSc surface Brillouin spectrometer (manufactured by Bede Scientific Instruments Ltd, Durham).

is a relatively weak process, so that the intensity of the inelastically scattered light of interest is rather weak compared with the elastically scattered light. Moreover, the frequency shift is typically 10–20 GHz, which is a very small fractional change in the wavelength of the light. The interferometer must therefore be of a very demanding specification. In this instrument it is a seven-pass parallel plate Fabry–Perot interferometer, with active feedback control of the interferometer spacing. The light passes to a detector with a low dark count, with full computer control of the whole instrument and the data acquisition.

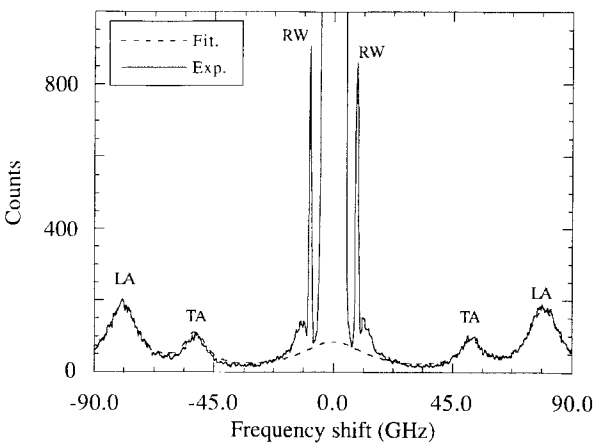
Brillouin spectra of silicon and of gallium arsenide are shown in Fig. 15. In both spectra peaks are present corresponding to bulk transverse and longitudinal waves, negative and positive frequencies corresponding as usual to Stokes and anti-Stokes scattering. These peaks are useful in bulk Brillouin spectroscopy, but are usually disregarded in surface Brillouin spectroscopy. The peaks corresponding to Rayleigh phonons are at lower frequencies, closer to the central elastic peak, at about 20 GHz for silicon and about 9 GHz for gallium arsenide. The particular values depend on the angle of incidence and the surface orientation.

An important application of surface Brillouin spectroscopy is in the characterization of subsurface damage in low-ductility materials due to mechanical surface engineering processes. Two examples from studies of gallium arsenide are shown in Fig. 16. The spectra were taken from GaAs (001) wafers, with propagation in a ⟨100⟩ direction. The anti-Stokes peaks are shown, corresponding to the frequency range 7.5–10.5 GHz of Fig. 15b. Figure 16a contains spectra from samples given a chemical polish and then polished with a series of alumina powders in sodium hypochlorate to give increasing mechanical damage. Figure 16b shows a chemically polished sample compared with two others polished using peroxide alkaline or deionized water as a lubricant. In each case increased subsurface damage leads to a smaller frequency shift, corresponding to a lower Rayleigh velocity, and also a broader peak, corresponding to higher attenuation.

From surface Brillouin spectra the velocity (and in principle the attenuation also) of surface acoustic waves can be deduced directly. Such measurements can then be



(a)



(b)

Figure 15. Brillouin spectra of: (a) Si(001) along ⟨110⟩; (b) GaAs(001) along ⟨110⟩; the angle of incidence was 60°. Horizontal axis is the frequency shift of the scattered light wave relative to the incident light. The vertical axis is the number of photon counts per data point. The peaks are identified as: LA, longitudinal acoustic phonon; TA, transverse acoustic phonon; RW, Rayleigh wave phonon. The broken line gives the best fit of the theoretical background spectrum (Courtesy of Dr P. Mutti).

used and analyzed in exactly the same way as measurements by quantitative acoustic microscopy. A detailed comparison of the two techniques is given in Table 7.2 of Ref. [63]. Surface Brillouin spectroscopy is not yet as accurate as quantitative acoustic microscopy, typically giving about 1% accuracy in surface wave velocity, but its much greater sensitivity to elastic changes within $\leq 0.3\,\mu m$ of the surface enables it to give excellent characterization of elastic properties in thin surface layers.

1.7 Time-Resolved Measurements

The quantitative techniques described so far have involved measuring, whether by $V(z)$ or by $V(f)$, modes propagating parallel to the surface of the specimen. If there are only one or two modes, this works well, but if there were many more then the analysis would quickly become unmanageable. It may then be useful to

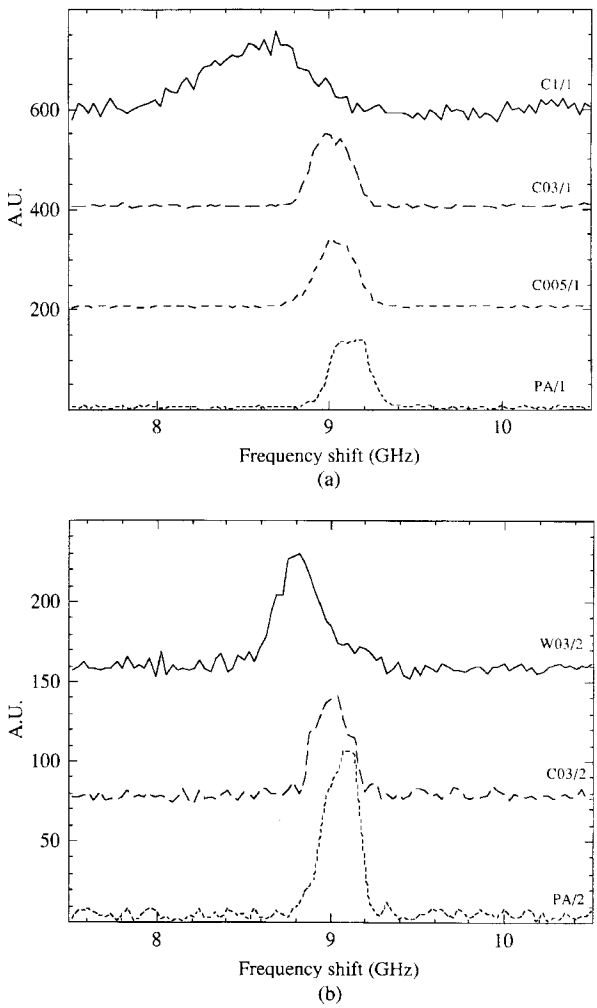


Figure 16. Brillouin spectra of GaAs(001) wafers with different surface preparations, showing the Rayleigh ⟨100⟩ propagation anti-Stokes peak. (a) Chemlox polished samples: these were lapped with 3 μm alumina powder on a glass plate and then chemically polished in a solution of peroxide alkaline (PA/1) to remove residual mechanical damage. Three samples were then polished with alumina powder of sizes 0.05 μm (C005/1), 0.3 μm (C03/1) and 1 μm (C1/1) in a solution of sodium hypochlorite (Logitech Chemlox procedure) on a soft pad using a Logitech PM2A polishing machine [63]. (b) Chemo-mechanically polished samples: one of the samples was PA polished (PA/2); the other two were polished with 0.3 μm alumina powder on a soft pad using respectively peroxide alkaline (C03/2) and deionized water (W03/2) as a lubricant [63].

consider measurements over a wide frequency range in a much more conventional way, viz. using a broad bandwidth to enable short pulses to be generated and measured. Using very short pulses, reflection from the top and bottom of a layer on a specimen can often be time-resolved.

An example of the kind of signal that can be used is shown in Fig. 17 [64]. The specimen was a coating of PET (polyethylene tetraphthalate) 15 μm thick on a stone-finish rolled steel substrate. Although there is some overlap of the two echoes, there is no difficulty in distinguishing them in order to measure the time difference. Since the shapes of the two pulses in Fig. 17 are so similar, a simple pattern-matching algorithm may be used, with great advantages in computational speed over Fourier domain techniques such as correlation or cepstral analysis. After suitable filtering, the turning points (i.e., the maxima and minima) of the waveform are found, and all the sets of three

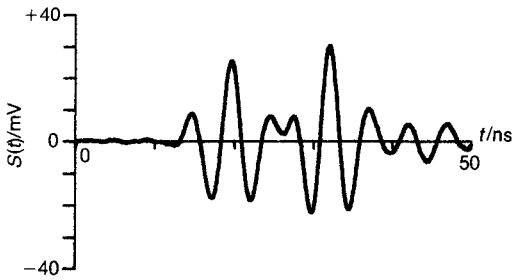


Figure 17. Signals reflected from the top and bottom surfaces of a 15 μm thick coating of polyethylene tetraphthalate on a stone-finish rolled steel substrate, using a short pulse of center frequency 230 MHz and half-power bandwidth 110 MHz; $z = +40$ μm, with the top surface of the polymer as the datum [64].

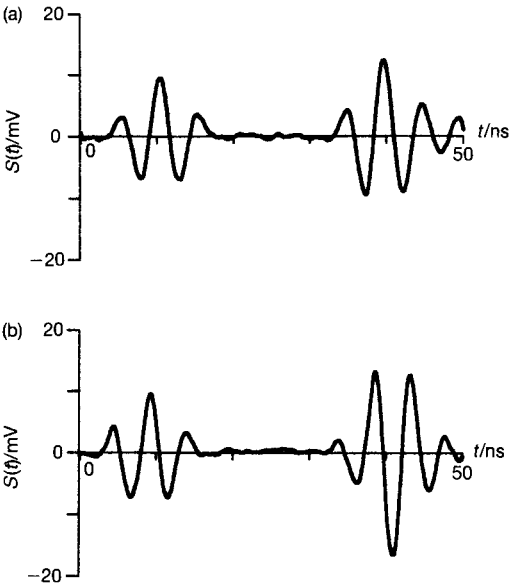


Figure 18. Signals reflected from the top and bottom surfaces of a 40 μm thick layer of polypropylene, with the same acoustic parameters as Fig. 10: (a) bonded to steel; (b) unbonded, with air backing [64].

adjacent points are identified satisfying the criteria of each point exceeding a minimum magnitude, and each adjacent pair exceeding a minimum separation in time and having second derivatives of opposite sign. For each set of three turning points, indicated by suffices a, b, c, at time t with size s, a figure of merit is calculated in order to find the set with best symmetry in time and size. The time asymmetry is defined as

$$A_t \equiv \frac{(t_c - t_b) - (t_b - t_a)}{t_c - t_a} \tag{34}$$

and the size asymmetry is defined as

$$A_s \equiv \frac{s_c - s_a}{s_b} \tag{35}$$

A figure of merit may then be defined as

$$F_A \equiv |s_b|(1 - \gamma|A_t| - \delta|A_s|) \tag{36}$$

in which γ and δ may be chosen empirically to give the best reliability of pulse identification; from experience the values $\gamma = 1$ and $\delta = 2$ work quite well. The positions of the pulses are determined from the two sets of three points with the greatest figures of merit. Hence the time difference can be measured; if required, the amplitudes of the two reflections can be determined from the corresponding values of s_b.

This method not only gives a computationally fast way of finding the thickness, it also enables disbonds between the coating and the substrate to be readily detected. In Fig. 18, signals from two films of polypropylene, each 40 μm thick, are shown. The first was adhered to a steel substrate (a), while the second was freely supported with air backing (b). Not only is the bottom surface reflection in the unbonded case larger than in the bonded case, but it has suffered a reverse in polarity. This is readily detected by the pattern matching algorithm, from the sign of s_b for the second echo.

If pulses can be generated and detected whose length is short compared with the time difference between reflections from

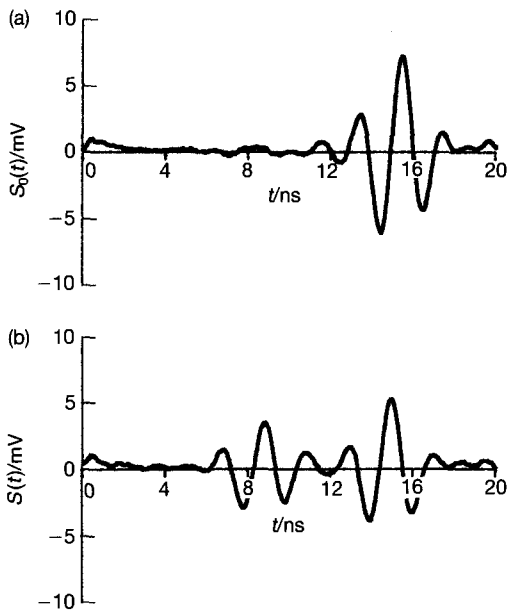


Figure 19. Short pulses for time-resolved measurements, the usable bandwidth of the pulses above noise level is about 0.5 GHz; they were digitized with an overall timing precision of 0.15 ns: (a) reference signal reflected from a glass slide at focus with no specimen; (b) reflected signal from a cell on the glass slide, with echoes from the top of the cell and from the interface between the cell and the substrate [66].

the top and the bottom surfaces of a layer, then the elastic properties of the layer can be deduced from the amplitude and timing of the two echoes [65]. The return pulses from such a situation are illustrated in Fig. 19a [66]. The first figure, Fig. 19a, is an oscilloscope trace of the reference echo from the substrate at defocus z_0 and with nothing on it except the coupling fluid. We can choose to write the reference signal as

$$S_0(t) \equiv A_0 s(t - t_0) \otimes g(t, z_0) \qquad (37)$$

The LHS represents what is actually measured in the experiment, and the RHS represents an interpretation in terms of the quantities of interest. A_0 is the reflection coefficient at the water–substrate interface; $s(t)$ is the two-way lens response ignoring any effects of focusing or attenuation in the coupling fluid; $g(t, z)$ describes the distortion of the wave shape due to defocus and attenuation. The symbol $\otimes$ denotes the convolution operation; in this context it enables the primary waveform to be described by $s(t)$, with adjustments being made by the function $g(t, z)$. Both these functions are experimentally determined. The waveform distortion function $g(t, z)$ may be arbitrarily set to unity at $z = 0$; the lens response function may then be measured with the substrate at the focus of the lens. In practice, if the defocus is positive (i.e., $z > 0$), the shape of the waveform is almost independent of defocus over the relevant range of z, so that g becomes a function of z only. Since g is now independent of t, the convolution in Eq. (37) reduces to a simple multiplication by a constant determined by the value of z.

If now the lens is moved to a region of the substrate that is covered by the layer of interest, two echoes will be received, one from the top surface of the layer, denoted by subscript 1, and one from the interface between the layer and the substrate, denoted by subscript 2. This is illustrated in Fig. 19b which was measured from a fibroblast cell on a polystyrene substrate. This signal can be written

$$S(t) \equiv A_1 s(t - t_1) \otimes g(t, z_1) + \\ + A_2 s(t - t_2) \otimes g(t, z_2) \qquad (38)$$

As in the case of the reference curve, provided the defocus is positive, it is adequate to allow g to be independent of t and to be a real function of z only, having been previously measured for the lens and substrate being used. The optimum value of z must be found experimentally, by scanning

through z and finding the minimum positive value at which the shape of the waveform remains approximately constant as a function of z. The amplitude may vary, but that does not matter, since it is allowed for in the dependence of g upon t. If the difference between t_1 and t_2 is greater than the length of the pulse, as is the case in Fig. 19b, then the two signals, from the top and the bottom surfaces, can be measured by calculating the normalized correlation of $S_0(t)$ and $S(t)$

$$C(t) = \frac{\int_{-\infty}^{\infty} S(t') \times S_0(t'+t)\,\mathrm{d}t'}{\int_{-\infty}^{\infty} S_0(t')^2\,\mathrm{d}t'} \tag{39}$$

In practice the limits of the integrals are the range of available data. There should be two peaks in the correlation, corresponding to the optimum match between the reference signals $S_0(t)$ and the two echoes contained in the signal $S(t)$. From the height and position of each maximum, four crucial parameters can be measured, namely $t_0 - t_1$, $t_0 - t_2$, A_1/A_0, A_2/A_0. Knowing the velocity v_0, impedance Z_0, and attenuation (taken as an average over the bandwidth) α_0, of the coupling fluid, and the impedance Z_s, of the substrate, all the acoustic properties of the layer can be determined; these are denoted by subscript 1.

(1) From the difference in time between the reference signal t_0 and the reflection from the top of the layer t_1, and knowing the velocity v_0 in the fluid, the thickness of the layer is

$$d = \tfrac{1}{2}(t_0 - t_1)v_0 \tag{40}$$

(2) From the time t_2 of the echo from the interface between the layer and the substrate, and the times of the other two echoes, the acoustic velocity in the layer is

$$v_1 = v_0 \frac{t_0 - t_1}{t_2 - t_1} \tag{41}$$

(3) From the ratio of the magnitude of the reflection A_1 from the top of the layer to the magnitude of the reference signal A_0, and knowing the impedance Z_0 of the coupling fluid and the impedance Z_s of the substrate, the impedance of the cell is

$$Z_1 = Z_0 \frac{A_0 + A_1}{A_0 - A_1} \tag{42}$$

(4) From the measurements of velocity and impedance, the density is immediately

$$\varrho_1 = \frac{Z_1}{v_1} \tag{43}$$

(5) Finally, from the amplitude A_2 of the echo from the interface between the layer and the substrate [which is described by an equation similar to Eq. (42)], and the amplitudes of the other two echoes, the attenuation in the cell is

$$\alpha_1 = \alpha_0 + \frac{1}{2d} \log_e\left(\frac{A_0}{A_2}\frac{Z_s - Z_1}{Z_s + Z_1} \times \frac{4Z_cZ_0}{(Z_c + Z_0)^2}\frac{Z_s - Z_0}{Z_s + Z_0}\right) \tag{44}$$

When two signals are so close together that they are not adequately separated in the correlation of Eq. (39), then it is better to express the signals in the frequency domain and work with their Fourier transforms, denoted by a bar. Assuming that the pulse shapes are not dependent on defocus, so that the frequency dependence of $\bar{g}(z)$ can be neglected, and again letting the LHS correspond to what is measured and the RHS represent this in terms of the

quantities of interest, the Fourier transforms of S_0 and S may be written

$$\bar{S}_0(f) = A_0\bar{s}(f)g(z_0)\exp(\mathrm{i}2\pi f t_0) \qquad (45)$$

and

$$\bar{S}(f) = A_1\bar{s}(f)g(z_1)\exp(\mathrm{i}2\pi f t_1) + \\ + A_2\bar{s}(f)g(z_2)\exp(\mathrm{i}2\pi f t_2) \qquad (46)$$

The simplest way to sharpen up such data in the frequency domain is to use a pseudo-inverse filter [67, 68]. The signals in the time domain can be thought of as a Dirac delta function $\delta(t - t_m)$ with amplitude A_m convolved with the lens time-response $s(t)$. The information that is needed is the values of t_m and A_m, which could be obtained by deconvolving the lens time-response. In a perfect world the Fourier transform of the measured signal $\bar{S}(f)$ divided by the reference signal $\bar{S}_0(f)$ would give exactly what is wanted, by the convolution theorem, but because it is inevitable that noise is present this might result in sometimes dividing by zeros, and in any case it would lead to wild distortions at frequencies where the reference signal is small. Therefore both signals are multiplied by the complex conjugate of the reference signal, to ensure that the denominator is real, and then a real number is added to the denominator to ensure good behavior when it is close to or below the noise level. The filtered signal is then, with complex conjugate denoted by *,

$$\bar{S}_2(f) = \frac{\bar{S}(f)\bar{S}_0^*(f)}{\bar{S}_0(f)\bar{S}_0^*(f) + M} \qquad (47)$$

The filtered signal is inverse transformed to give a shorter pulse in the time domain, and it can then be analyzed by the correlation of Eq. (39).

A more powerful technique for analyzing the Fourier transformed signals begins with cepstral filtering [69]. The logarithm is taken of the modulus of each of the two equations, and the first is subtracted from the second to give

$$\ln|\bar{S}(f)| - \ln|\bar{S}_0(f)| = \\ = \ln|A_1\bar{s}(f)g(z_1)\exp(\mathrm{i}2\pi f t_1) + \\ + A_2\bar{s}(f)g(z_2)\exp(\mathrm{i}2\pi f t_2)| - \\ - \ln|A_0\bar{s}(f)g(z_0)\exp(\mathrm{i}2\pi f t_0)| \qquad (48)$$

The terms "cepstrum" and "cepstral" come from inverting the first half of the words spectrum and spectral; they were coined because often in cepstral analysis one treats data in the frequency domain as though it were in the time domain, and vice versa. The value of cepstral analysis comes from the observation that the logarithm of the power spectrum of a signal consisting of two echoes has an additive periodic component due to the presence of the two echoes, and therefore the Fourier transform of the logarithm of the power spectrum exhibits a peak at the time interval between them. The additive component in the logarithm of the power spectrum comes from a multiplicative component in the power spectrum itself, just as the subtraction of the logarithms in Eq. (48) corresponds to the division in Eq. (47). Figure 20a is a cepstrum of the signal in Fig. 19b.

If several oscillations in $\ln|\bar{S}(f)| - \ln|\bar{S}_0(f)|$ occur over the frequency range of measurement, as in Fig. 20a, then the period Δf of the oscillations can be measured by a Fourier transform into the time-interval domain. This transform cannot be related to absolute time because information relative to t_0 has been lost, but it will be possible to identify a peak corresponding to $t_2 - t_1 = 2d/v_c$. When, however,

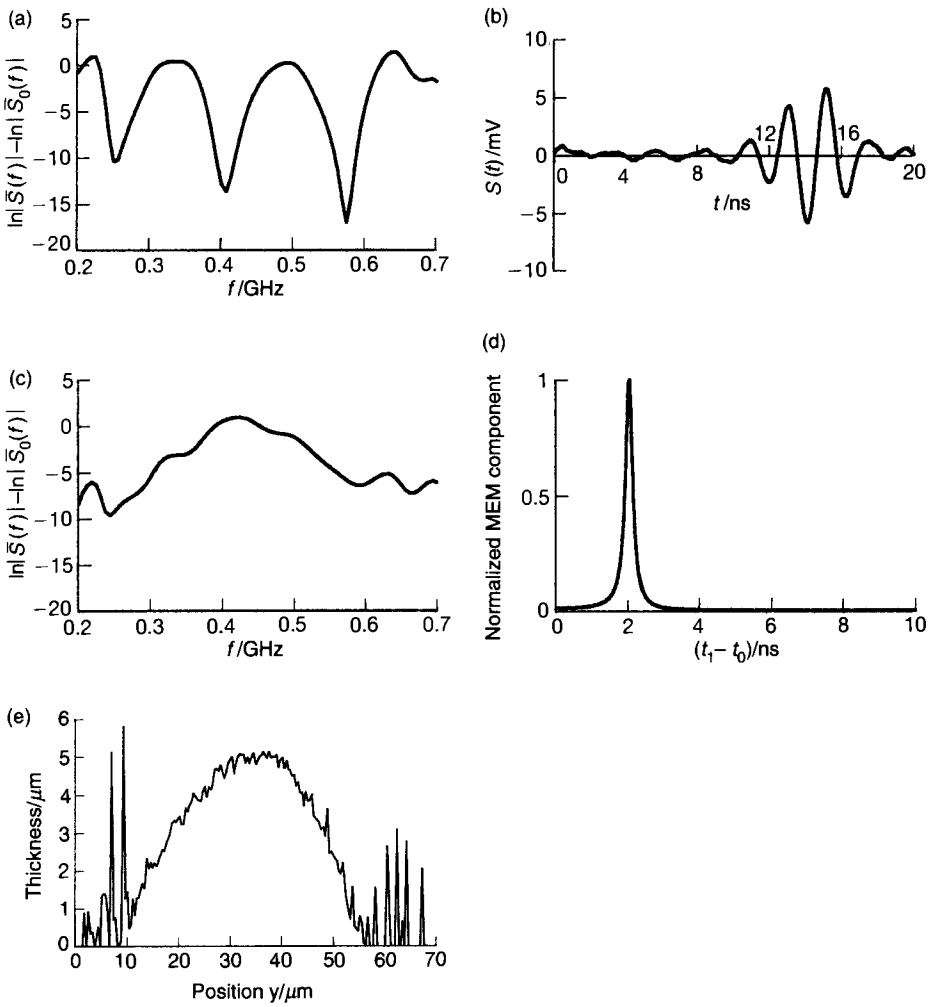


Figure 20. Cepstral and maximum entropy (MEM) analysis of time-resolved signals: (a) The cepstral analysis function $\ln|S(f)| - \ln|S_0(f)|$ in the frequency domain for the data in Fig. 19, using equation (48) with the Fourier transforms of signals in Fig. 19a and b as $S_0(f)$ and $S(f)$ respectively. (b) Two overlapping and unresolved signals in the time domain from a very thin part of a cell on a glass substrate. (c) The cepstral analysis function $\ln|S(f)| - \ln|S_0(f)|$ in the frequency domain for the signal in (b). (d) The result in the time-interval domain of MEM analysis of the cepstral function in (c). (e) Thickness of a cell deduced from cepstral and MEM analysis of signals measured along a line. The horizontal and vertical scales are not the same, and the thickness measurement becomes unreliable below 1.5 μm [66].

less than one oscillation is present in $\ln|\bar{S}(f)| - \ln|\bar{S}_0(f)|$, the value of $t_2 - t_1$ cannot be found simply, because this corresponds to the value of $t_2 - t_1$ being less than the usable bandwidth β of the transducer, or equivalently to the pulses overlapping in time. Of course, if $t_2 - t_1 \ll 1/\beta$, then there will come a point beyond which no amount of signal processing will recover the information, but in mild cases it is possible to achieve remarkable measurements from otherwise

unusable data using the maximum entropy method, or MEM [67]. MEM may be thought of as an alternative operation to the Fourier transform, which transforms $\ln|\bar{S}(f)| - \ln|\bar{S}_0(f)|$ back into the time-interval domain. The reason that a Fourier transform does not give a very useful result when there are not well developed oscillations in $\ln|\bar{S}(f)| - \ln|\bar{S}_0(f)|$ is that information is not available over a sufficiently large range of f. MEM compensates for this by finding the transform that would correspond to the maximum possible number of different shapes of curve outside the range of f for which data is available. The term maximum entropy comes from the analogy with statistical thermodynamics, where entropy is a measure of the logarithm of the number of different arrangements that give the same overall state.

An example of the application of the power of cepstral analysis and MEM is illustrated in Figs. 20b to e [66]. The signal in Fig. 20b was taken from a specimen similar to the one in Fig. 19b, but closer to the edge where the cell was thinner. The overlap of the signals is so great that there is no hope of separating them by eye. The application of cepstral analysis gives the curve shown in Fig. 20c, and there is not even one full relevant oscillation in the frequency domain. But the maximum entropy method enables useful information to be obtained even from such poorly resolved data as this, and in the time-interval domain in Fig. 20d the MEM transform of $\ln|\bar{S}(f)| - \ln|\bar{S}_0(f)|$ has pronounced peak from which $2d/v_1$ for the cell at that point can be determined. The time separation is about 2 ns, corresponding to a thickness of less than 2 μm. This may be a world record for acoustic distance resolution in this way. Lest anyone should think that it is simply coincidence that Fig. 20d shows a peak at the reciprocal of the bandwidth in Fig. 20a and c, Fig. 20e is a plot of the thickness of the cell along a diameter, measured directly from the position of the peak in a series of curves like the one in Fig. 20d. A reasonable profile is obtained over a range of thickness from 5 μm at the center to about 1.5 μm near the edge.

1.8 Crack Depth

One of the most remarkable, and most challenging, applications of time resolved acoustic microscopy is in the measurement of the depth of short cracks [70]. The measurement of the time-of-flight of an ultrasonic signal has been established for some time as an accurate way to measure the size of cracks in structures [71, 72]. The size needs to be known accurately in order that fracture mechanics may be used to calculate the growth behavior of a crack, and hence the integrity for service may be evaluated. Much conventional ultrasonic nondestructive testing uses specular reflections from defects. Using signal averaging techniques, it is possible to detect waves that have been diffracted from a crack tip with sufficient sensitivity to measure the time-of-flight of their propagation. This can be much more accurate and reliable than simply looking for a probe position and orientation at which the signal drops to, say, half its maximum value. Often the time-of-flight technique is used with a combination of two probes, one to send and one to receive. In that case a single measurement is ambiguous; it gives a locus in the shape of an ellipse on which the

crack tip must lie, with the probes at the two foci. A series of measurements must be made, moving the two transducers together over the surface. If the result is plotted in a suitable form, then by pattern recognition is it possible to identify the true position of a crack tip.

Because Rayleigh waves can be excited so strongly in the acoustic microscope, the easiest place to start looking for crack tip diffraction is where a crack edge meets the surface [73]. Figure 21 illustrates this. In Fig. 21, the horizontal axis is time. Each line is like an oscilloscope trace, but with the value of the signal at each instant represented by the brightness of the spot rather than by a vertical deflection. A positive signal is white, a negative signal is black, and a zero signal is mid grey. Thus short pulses such as the ones in Figs. 17 to 19 would appear as a series of bright and dark bands. The vertical axis in Fig. 21 denotes lens movement in the y-direction, parallel to the surface of the specimen and perpendicular to the line along which the crack meets the surface. In Fig. 21a the first strong band is the specualar reflection from the surface, and the second strong band is the Rayleigh reflection, in which rays at the Rayleigh angle in the fluid excite Rayleigh waves in the surface of the specimen, which in turn generate waves in the fluid which return to the lens (cf. Fig. 5a). If the surface were uniform, all the reflections in such plots would be vertically uniform bands. But the presence of the crack leads to an extra $\times$ pattern, whose crossing point occurs when the lens is directly over the crack mouth. As the lens approaches the crack, a Rayleigh wave is reflected at the crack which returns to the far side of the lens, and therefore gives a signal which arrives later than the main Rayleigh signal. Once the lens has moved sufficiently close to the crack so that Rayleigh waves can be excited on both sides of it, reflections can appear before and after the

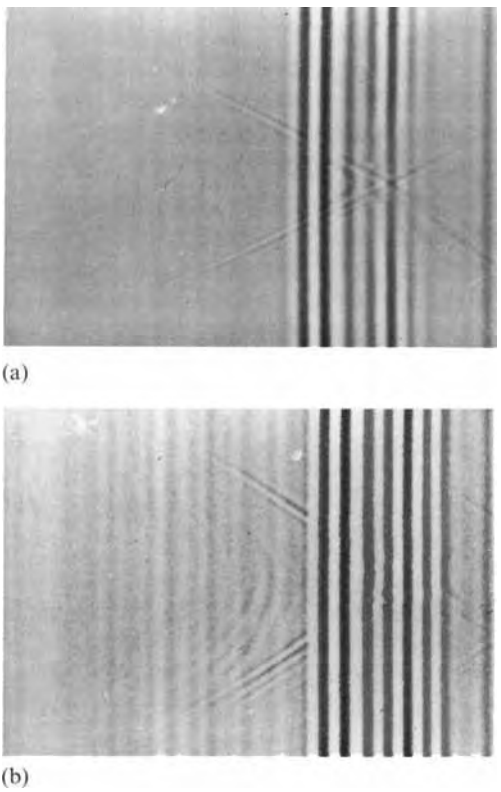


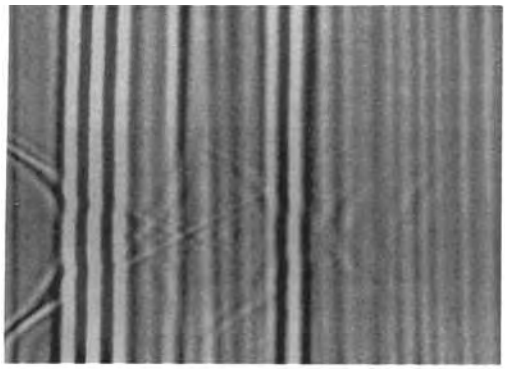
Figure 21. Time-resolved $S(t, y)$ scan along a line perpendicular to a crack in glass, scanning across the crack: (a) some distance from the end of the crack; (b) 30 μm from the end of the crack. The horizontal axis is time t; the vertical axis is y, and the value of the signal S is indicated by the intensity, with mid-grey as zero and dark and light denoting negative and positive values, respectively. In both figures, the first echo (seen as the first stripy vertical bar) is the geometric reflection from the surface of the specimen, and the second echo (seen as the second stripy vertical bar) is the Rayleigh reflection. The patterns forming a $\times$ are the reflections from the near and the far sides of the crack, which cross over when the lens is directly above the crack. In (b), where the scan passes quite near to the tip of the crack, the hyperbolic pattern is due to the crack-tip-diffracted wave [73].

main Rayleigh reflection. As the lens approaches a position overhead of the crack, these three coincide, separating again on the other side of the crack. All this can be seen in Fig. 21a. Figure 21b shows what can happen when the lens crosses the crack close to its tip. Now there is an extra set of echoes, following an approximately hyperbolic pattern. These are due to Rayleigh waves excited in the surface that are diffracted from the crack tip itself. They show that the Rayleigh wave crack-tip-diffracted signal is strong enough to be detected and measured in the acoustic microscope.

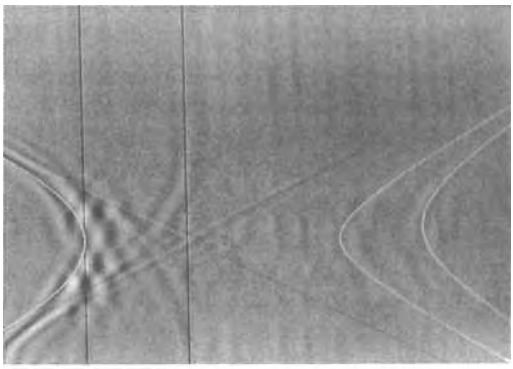
Of course, the true challenge is not to measure the position of a crack tip at the surface, but inside a specimen [74]. To investigate this, measurements like those of Fig. 21 were made of transparent specimens of polystyrene containing shallow cracks. A ray theory was developed to calculate the arrival time of signals which traveled along all the paths that were considered to be sigificant. A field theory was also developed, to check the ray theory and to indicate the relative amplitudes of the different signals. Then the results of the ray theory were fitted to the experimental data, taking the position of the crack tip (both its depth and also any lateral displacement relative to the crack mouth) as the unknown to be fitted. Finally, the position of the crack tip was measured optically using a standard light microscope. In a series of measurements of crack depths from 50 to 75 μm, the two kinds of measurement gave the same depth to better than 5%.

Figure 22a shows measurements made near the end of a crack in polystyrene. In polymers, lateral longitudinal waves replace Rayleigh waves, because the shear wave velocity is too slow. The specular and lateral wave reflections from the surface rather dominate the other signals, so in Fig. 22b it is possible to see a curved pattern on the left similar to the crack-tip-diffracted signal in Fig. 21b. There are also two curved patterns which merge into the lateral wave reflection away from the crack mouth; the curved patterns are due to waves excited in the fluid by lateral waves incident on the crack, or vice versa [75]. But the important patterns are the two hyperbola-like ones on the right. They are due to bulk waves diffracted from the crack tip. The fact that there are two means that there are two diffracting features. Their positions were deduced by fitting ray theory curves to the data, and the results are illustrated in Fig. 22c. Subsequent examination in an optical microscope confirmed that there was a kink near the bottom of the crack, and that the acoustic microscope was detecting the kink as well as the tip. The comparison between the acoustic and the optical measurements is very satisfactory, as is the depth discrimination between the kink and the tip. Figure 22d shows the ray theory curves calculated from the optical data superimposed on the acoustic data. There are no free parameters. The second white curve on the right is displaced asymmetrically from the first because the tip is displaced sideways relative to the kink. These experiments in a transparent model material give some confidence for measurements inside a metal.

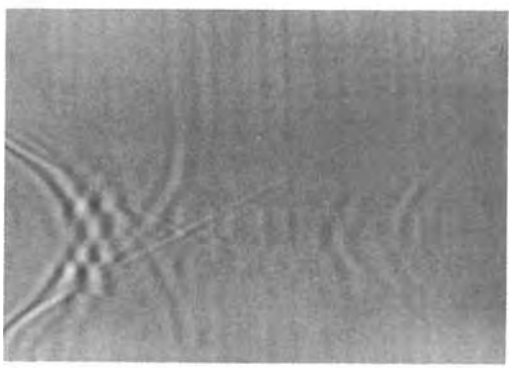
Crack depth measurements in a metal are more difficult than in a polymer. The acoustic impedance and velocity are greater, and therefore less energy is transmitted into the solid, and what does enter



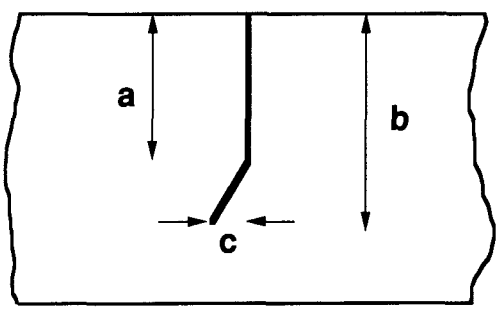
(a)



(d)



(b)



μm	a	b	c
optical	52.5	75.5	10
acoustic	54.5	72	9

(c)

Figure 22. Crack in polystyrene: (a) time-resolved scan; (b) time-resolved scan after image processing to remove specular and Rayleigh reflection; (c) schematic of section through crack; (d) ray calculations based on optical measurements superimposed on (b); $z = -60\,\mu m$ [74].

can excite both longitudinal and shear waves. But good signal averaging and processing can be used to detect sufficient signal strength, and the ray model can be extended to allow for the additional wave types. Figure 23a is an acoustic micrograph of a short fatigue crack in an aluminum lithium 8090 alloy [74]. Aluminum-based alloys have low crystal elastic anisotropy, and therefore low grain scattering of acoustic waves. A time-resolved scan was made along the white line, and this is shown after image processing in Fig. 23b. The position corresponding to the vertex of the shear wave diffracted into a shear wave at the crack tip is indicated by a white arrow. This corresponds to a crack 42 μm deep, with a lateral displacement between the tip and the mouth of 7 μm. After these measurements the specimen was sectioned, and an acoustic micrograph of the section corresponding to the white line in Fig. 23a is shown in Fig. 23c. Measurements of this picture indicate a crack depth of 43 μm with a

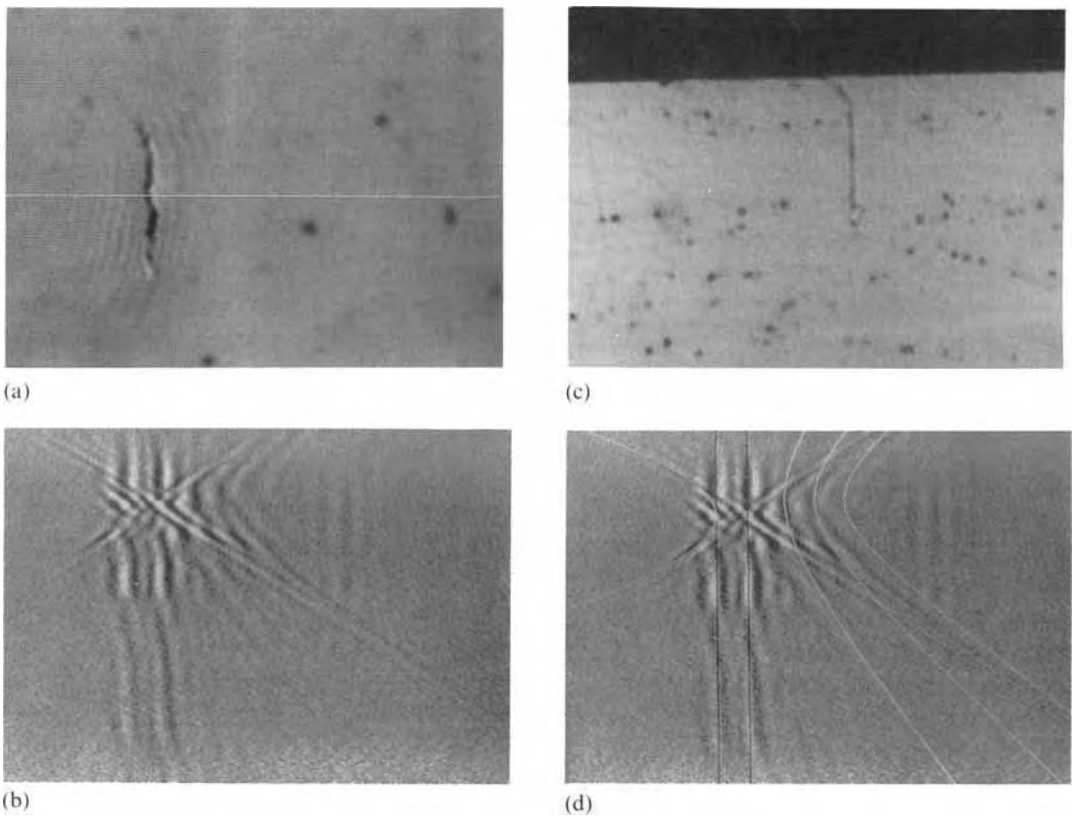


Figure 23. Fatigue crack in Al–Li 8090 alloy: (a) acoustic image, 350 MHz, $z = 0$; (b) time-resolved scan along line in (a), after image processing, $z = -40\,\mu m$; (c) acoustic image of cross section corresponding to line in (a), 500 MHz, $z = 0$; (d) ray calculations based on (c) superimposed on (b) [74].

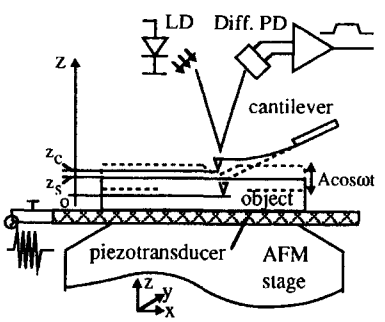
lateral displacement of 6.5 μm. The kink near the mouth is difficult to identify or measure in Fig. 23b, because of the much stronger specular and Rayleigh reflections. In Fig. 23d the ray theory curves calculated from the position of the tip in Fig. 23c are superimposed on Fig. 23b. Once again there are no free parameters. There are three diffracted curves from a single crack tip, corresponding to longitudinal–longitudinal, longitudinal–shear, shear–shear (in order of arrival), because there may be mode conversion from one to the other in the diffraction process at the tip.

1.9 Ultrasonic Force Microscopy

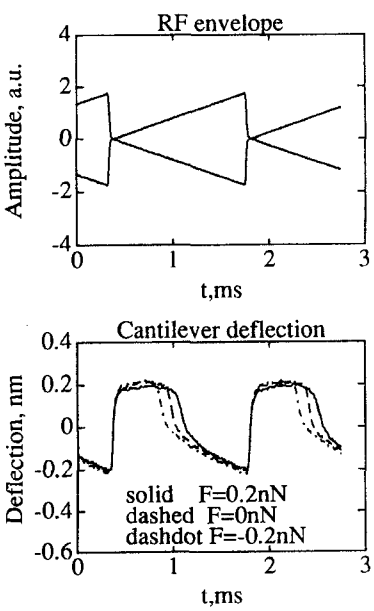
How can one go beyond the limitation of the resolution of 0.7 μm or so imposed by the attenuation in the coupling fluid? Attempts to use cryogenic fluids with lower attenuation have been successful in achieving higher resolution, but at the expense of the unique information about the elastic properties of the sample, so that there is no advantage over scanning probe techniques that give topographic

information with higher resolution still. We have seen in Section 6 of this Chapter that higher frequency Rayleigh waves can be generated and measured using surface Brillouin spectroscopy, but that technique is generally used in applications where the lateral resolution (from diffraction-limited light) is not an issue. Dramatic advances in the detection of surface elastic properties with high spatial resolution have come through the development of ultrasonic force microscopy.

The ultrasonic force microscope (UFM) consists of an atomic force microscope with the sample mounted on an ultrasonic transducer [76]. The principle is illustrated in Fig. 24. Ultrasonic waves propagate through the sample, and cause the surface of the sample to vibrate with displacement $A\cos\omega t$ (Fig. 24a). The cantilever of the atomic force microscope is used to detect the surface vibration at each point. It would be possible in principle to detect the surface vibrations at the ultrasonic frequency itself, but because these are generally well above the resonant frequency of the cantilever, the response of the cantilever at megahertz ultrasonic frequencies tends to be rather weak. A much more ingenious detection scheme exploits the nonlinear nature of the contact between the cantilever tip and the sample



a)



b)

Figure 24. The principle of an ultrasonic force microscope (UFM). (a) A piezoelectric transducer is mounted on the sample stage of an atomic force microscope, and the sample is placed in contact with the transducer. A radio-frequency signal is applied to the transducer, causing the surface of the sample to vibrate with local displacement $A\cos\omega t$. The nonlinear contact between the sample surface and the cantilever tip causes a mean deflection z_c of the cantilever from its static position z_s. The cantilever deflection is measured by a beam of light from a light-emitting diode (LD) which is reflected by the surface of the cantilever and detected by a differential photodiode (Diff. PD). (b) The upper oscilloscope trace shows the modulation of the ultrasound in the UFM. The ultrasonic frequency was 3 MHz, the modulation frequency was 700 Hz, and the maximum voltage amplitude applied to the transducer was 1 V. The lower oscilloscope trace shows the response of the cantilever for three different values of applied static force, with negative deflection corresponding to movement away from the sample surface. The cantilever was microfabricated Si_3N_4 with a spring constant 0.024 M m^{-1} and free resonant frequency 33 kHz [76].

surface. At its simplest, this is a contact which can push but not pull; a slightly more sophisticated description is that the stiffness of the contact depends on the contact area and hence on the instantaneous force. Regardless of the particular details of the force displacement law, the fact that it is nonlinear means that when the surface vibrates, there will be a quasi-static (d.c.) displacement of the cantilever as well as the oscillating (a.c.) vibration at the ultrasonic frequency. It is difficult to measure a static displacement in an atomic force microscope because of thermal drift and also because when scanning it would be indistinguishable from topography, and so the amplitude of the ultrasonic vibration is modulated as some frequency at which the cantilever can easily respond; this has the further advantage that the cantilever deflection can then be readily measured using standard atomic force microscope detection electronics.

In this way the contact between the cantilever tip and the sample surface is used as a kind of mechanical diode to detect the modulation of the ultrasonic excitation, rather as an electronic diode in a crystal radio is used to detect amplitude modulation of the carrier waveform. The principle is demonstrated in Fig. 24b. The upper figure represents the oscilloscope trace of the envelope of the ultrasonic excitation. Triangular modulation is used, so that the ultrasonic amplitude increases steadily to some maximum value, before returning to zero and starting again. The lower figure represents the response of the cantilever (with negative voltage corresponding to repulsion of the tip away from the surface). The highly nonlinear response is evident. At low amplitudes scarcely any deflection is visible, but as the amplitude is increased a threshold is reached above which the deflection of the cantilever increases dramatically. The detected signal at the modulation frequency can be passed through a lock-in amplifier and used to modulate the brightness of an ultrasonic force image.

An ultrasonic force image is compared with other scanning force images in Fig. 25. The sample was a computer floppy disk, composed of 10–20 μm metal-coated ferrite particles in a relatively soft polymer binder. Figure 25a is an ultrasonic force image, obtained with an ultrasonic frequency of 6 MHz and a modulation frequency of 2.4 kHz. Figure 25b is a standard topography image obtained with constant normal force. It is immediately apparent from direct comparison that the ultrasonic image is giving information different from the topography image. Figure 25c is a force modulation mode image, in which the sample is vibrated at a frequency well below the resonant frequency of the cantilever. This is expected to give a greater response where the sample is stiffer. The ultrasonic image has some similarities with this, but it seems to give additional information, and also sharper interfaces between regions of different properties. The resolution which can be obtained in an ultrasonic force microscope is demonstrated in Fig. 26, which shows two images of mica. In Fig. 26a lattice resolution is obtained, and in 26b a monatomic step is imaged.

What property is revealed in an ultrasonic force image? At the simplest level the contrast is determined by the local stiffness of the material at the surface. This is similar to what is imaged in the force modulation mode. The contact and the

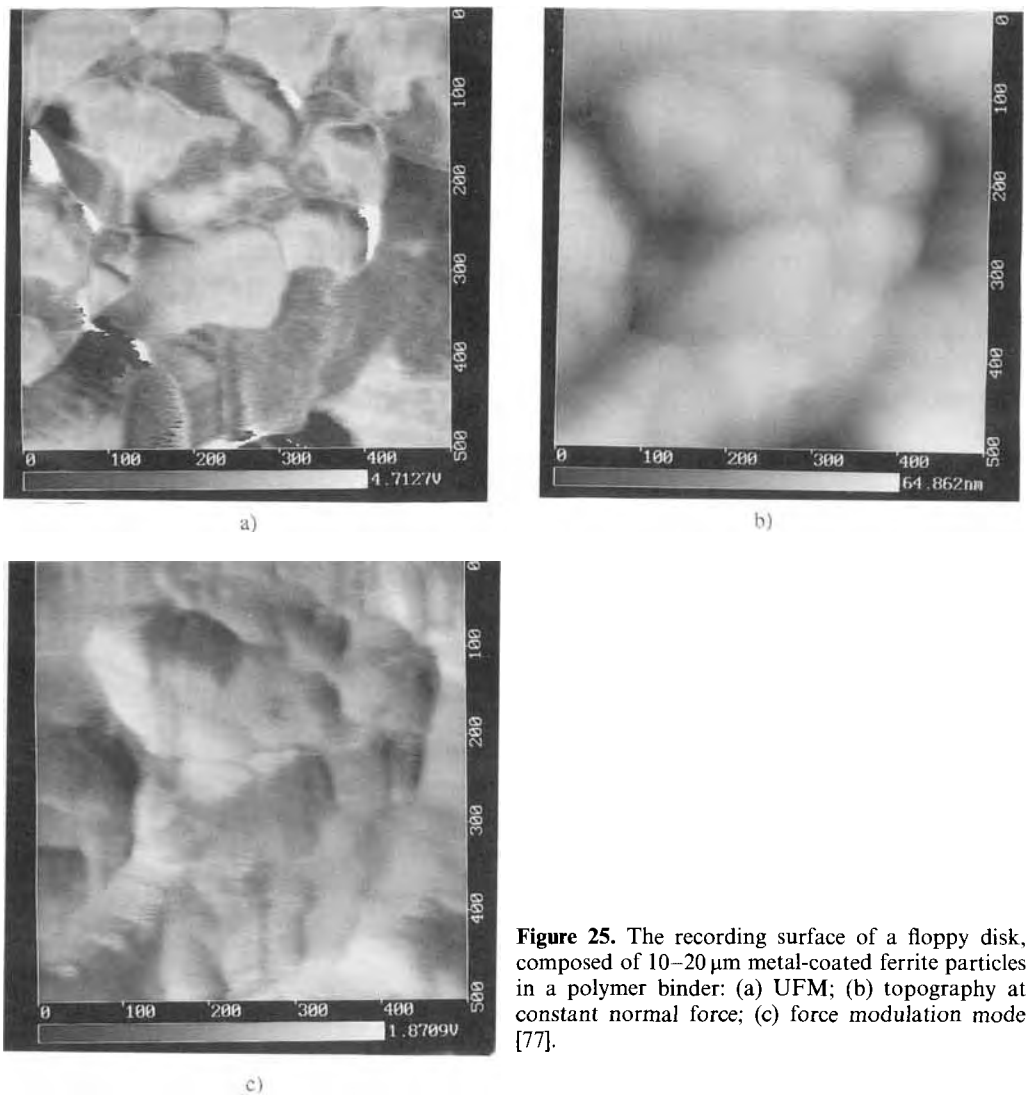


Figure 25. The recording surface of a floppy disk, composed of 10–20 μm metal-coated ferrite particles in a polymer binder: (a) UFM; (b) topography at constant normal force; (c) force modulation mode [77].

cantilever may be considered as two springs in series, with the effective mass of the cantilever tip between them. As the sample surface is moved up and down, so the mass will move also. At frequencies much below resonance, the amplitude of the movement of the mass is determined by the relative stiffness of the two springs. For most solid samples the stiffness of the sample-tip contact is much greater than the stiffness of the cantilever, and so the displacement amplitude of the tip is only slightly less than the amplitude of the surface vibrations. This gives rather weak contrast even for large relative variations of the sample stiffness. By increasing the frequency well above the resonant frequency of the cantilever, the contact also has to push against the inertia of the effective mass of the cantilever tip, and

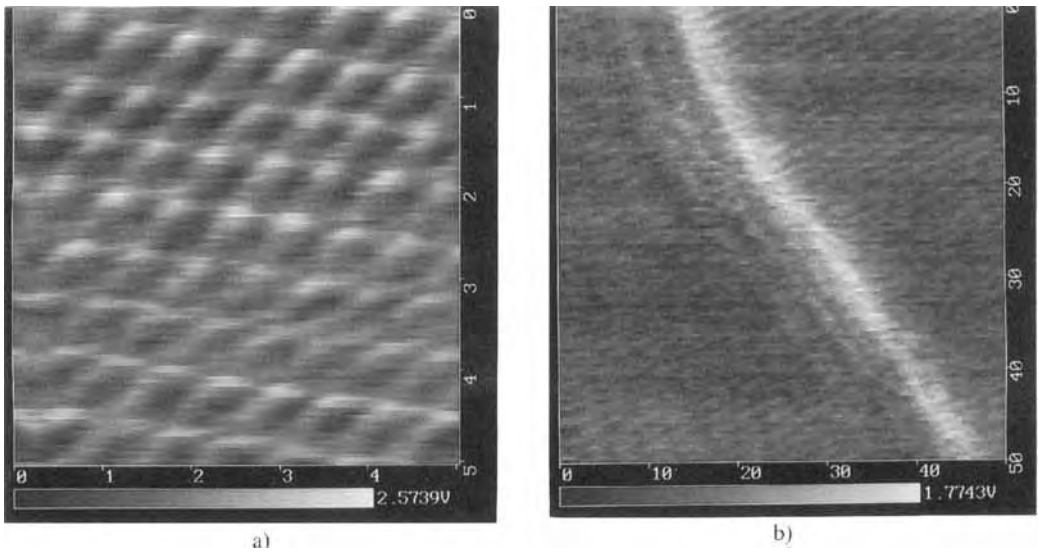


Figure 26. UFM images of mica: (a) lattice resolution; (b) monatomic step. The ultrasonic frequency was 6 MHz, the voltage amplitude applied to the transducer was 0.25 V, and the modulation frequency was 2.5 kHz [77].

this gives a response which is much more dependent on the local sample stiffness. A detailed account of the principle of the ultrasonic force microscope gives a useful indication of the contrast mechanism in the limit that the effective mass of the cantilever tip is infinite, so that only the quasi-static displacement is significant [78]. This explains, for example, why the threshold ultrasonic amplitude evident in Fig. 24b depends on the force that is applied to the cantilever.

If the bandwidth of the detection system is extended to include the ultrasonic frequency, then it proves possible to detect the displacement of the cantilever at that frequency [79]. Indeed, it is even possible to watch the propagation of the vibrations along the cantilever, which seems to occur as an antisymmetric plate mode at a velocity considerably less than a bulk ultrasonic velocity. Since the contact between the tip and sample is nonlinear, it might be expected that the system will exhibit the kind of phenomena associated with driven nonlinear oscillators. This proves to be the case [80]. As the ultrasonic amplitude is increased, there comes a point at which the downward acceleration of the surface cannot be sustained by the force of the cantilever on the tip, and contact can be lost. Depending on the amplitude, contact may be regained during the same ultrasonic cycle, or, if the tip has been thrown off with sufficient velocity, it may be regained in some subsequent cycle. Such orbits may be stable or unstable. The unstable orbits lead to chaotic motion of the tip, the stable orbits may be subharmonics of the ultrasonic frequency. An example in which the tip makes one vibration for every eight cycles of the ultrasonic frequency is shown in Fig. 27. The subharmonic behavior depends on the

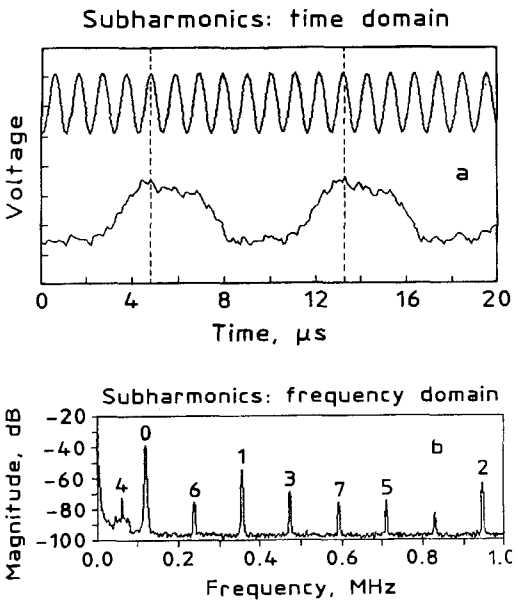


Figure 27. Nanosubharmonics in a UFM. (a) The upper oscilloscope trace shows the excitation of the transducer with a period of 1.06 ns, and the lower oscilloscope trace shows the response of the cantilever with a period of 8.48 ns, i.e., period 8 (the voltage scale of the oscilloscope is uncalibrated). (b) Fourier transform of the cantilever response displayed on a spectrum analyzer: the peaks are numbered according to their magnitude. Peak 2 is the excitation frequency, 948.75 kHz; peak 0 is the period 8 subharmonic, 117.5 kHz, which is 24 dB stronger; peak 4 is the free cantilever resonance; the other peaks are harmonics of peak 0 [80].

ratio of the surface acceleration amplitude to the acceleration that can be provided by the cantilever spring, and so by altering either the ultrasonic amplitude or the normal force it is possible to progress through a cascade of subharmonics separated by regions of chaos. The boundaries between different regimes are determined by details of the elastic contact, and also by the energy lost per bounce. In a large scale system this might be described by the coefficient of restitution, which would be largely determined by the bulk material's properties. At the nanoscale of the contact in an ultrasonic force microscope it is likely that adhesion plays a dominant role, energy being lost because the adhesive forces at a given separation may differ when the surfaces are coming together from when they are moving apart. Thus the phenomenon of nanosubharmonics offers the possibility of observing and imaging dynamic adhesion phenomena with the spatial resolution of an atomic force microscope.

Ultrasonic force microscopy makes it possible to image the dynamical mechanical properties of surfaces with nanometer resolution. In the diode detection mode, it gives images which have considerably more sensitive contrast, especially to stiffer materials, than an atomic force microscope operating in a lower frequency force modulation mode. In the nanosubharmonic mode, it is also possible to observe the effects of adhesion. Because the ultrasonic force uses near-field detection, there is no diffraction limit to the resolution that can be obtained. And because no fluid coupling is needed and the modulation frequency can be low, there is no limit to the frequency imposed by fluid attenuation or the need for short pulses. Ultrasonic force microscope images have been demonstrated at 114 MHz [78], and it may well prove possible to use even higher frequencies still.

1.10 Conclusion

The key message of acoustic microscopy is that it enables the interaction of acoustic waves with elastic properties and features to be visualized with microscopic

resolution. In solids a dominant role is often played by Rayleigh waves and other surface waves, and the dependence of their velocity on bulk properties and anisotropy and surface layers. There is especially strong sensitivity to scattering by surface cracks and boundaries. At lower resolutions, and in polymeric materials, Rayleigh waves are less dominant, and useful interior imaging is possible. This can bridge the gap between high resolution acoustic microscopy and conventional nondestructive testing, and the enhanced depth discrimination of a confocal system can be exploited. With thin coatings and biological cells the echoes from the top of the specimen and the interface with the substrate can be separated.

For diffraction limited surface imaging with water coupling the practical upper frequency is 2 GHz, giving a wavelength of 0.75 μm. For interior imaging lower frequencies are used, a frequency of 100 MHz giving a resolution of 50 μm in a typical solid. Nanosecond pulses at higher frequencies enable depth discrimination better than 5 μm to be achieved. For quantitative Rayleigh wave measurements of surfaces, a frequency of 225 MHz is standard, sampling a depth of 10–20 μm. These limits of depth sampled and resolution can be overcome in two ways. The depth sampled can be reduced by using higher frequencies excited by interaction with light. Using surface Brillouin spectroscopy depths less than 0.3 μm can be sampled by Rayleigh waves at frequencies up to 20 GHz, enabling quantitative measurements to be made of elastic properties in very thin surface layers. Using ultrasonic force microscopy, the diffraction limit can be exceeded and near-field acoustic images can be made with nanometer resolution and information about dynamic elastic properties and adhesion.

Acknowledgements

I am deeply grateful to all the colleagues and friends, in Oxford and internationally, who have contributed ideas and results. I am also grateful to Oxford University Press for allowing me to use figures from the monograph *Acoustic Microscopy* (1992), in which further details of many of the topics discussed here will be found, together with a more comprehensive description of the technique and its use for imaging as well as for quantitative measurements.

1.11 References

[1] G. A. D. Briggs, *Acoustic Microscopy*. Oxford: Clarendon Press **1992**.

[2] L. W. Kessler, Acoustic microscopy commentary: SLAM and SAM. *IEEE Trans. Sonics Ultrason.* **1985**, *SU-32*, 136–138.

[3] W. Arnold, Acoustic imaging in materials evaluation. *Europhys. News* **1991**, *22*, 171.

[4] G. S. Cargill, Electron beam acoustic imaging, in: *Physical Acoustics*, Vol. 18 (Eds. W. P. Mason, R. N. Thurston). San Diego: Academic **1988**.

[5] G. Gremaud, A. Kulik, S. Sathish, Scanning acoustic microscopy: A physicists tool. *Europhys. News* **1991**, *22*, 167–170.

[6] G. M. Crean, C. M. Flannery, S. C. O Mathuna, Acoustic microscopy of microelectronic interconnection and packaging technologies, in: *Advances in Acoustic Microscopy, 1* (Ed. G. A. D. Briggs). New York: Plenum, **1995** pp. 1–48.

[7a] G. Pfannschmidt, Characterization of electronic components by acoustic microscopy, in: *Advances in Acoustic Microscopy 2* (Eds. G. A. D. Briggs, W. Arnold). New York: Plenum **1996**.

[7b] J. Bereiter-Hahn, Probing biological cells and tissues with acoustic microscopy, in: *Advances in Acoustic Microscopy 1* (Ed. G. A. D. Briggs). New York: Plenum **1996**, 79-M5.

[8] T. Wilson, C. J. R. Sheppard, *Theory and practice of scanning optical microscopy*. London: Academic **1984**.

[9] K. Yamanaka, Y. Enomoto, Observation of surface cracks with scanning acoustic microscope. *J. Appl. Phys.* **1982,** *53*, 846–850.

[10] H. K. Wickramasinghe, Contrast in reflection acoustic microscopy. *Electron. Lett.* **1978,** *14*, 305–306.

[11] H. K. Wickramasinghe, Contrast and imaging performance in the scanning acoustic microscope. *J. Appl. Phys.* **1979,** *50*, 664–672.

[12] A. Atalar, An angular spectrum approach to contrast in reflection acoustic microscopy. *J. Appl. Phys.* **1978,** *49*, 5130.

[13] A. Atalar, A physical model for acoustic signatures. *J. Appl. Phys.* **1979,** *50*, 8237.

[14] C. F. Quate, A. Atalar, H. K. Wickramasinghe, Acoustic microscopy with mechanical scanning: a review. *Proc. IEEE* **1979,** *67*, 1092–1114. Reprinted (1986) in: *Modern Acoustical Imaging* (Eds. H. Lee, G. Wade). New York: IEEE, pp. 180–202.

[15] C. J. R. Sheppard, T. Wilson, Effects of high angles of convergence on *V(z)* in the scanning acoustic microscope. *Appl. Phys. Lett.* **1981,** *38*, 858–859.

[16] K. K. Liang, G. S. Kino, B. Khuri-Yakub, Material characterisation by the inversion of *V(z)*. *IEEE Trans. Sonics Ultrason.* **1985**, *SU-32*, 213–224.

[17] J. A. Hildebrand, K. Liang, S. D. Bennett, Fourier-transform approach to material characterization with the acoustic microscope. *J. Appl. Phys.* **1983,** *54*, 7016–7019.

[18] C. Ilett, M. G. Somekh, G. A. D. Briggs, Acoustic microscopy of elastic discontinuities. *Proc. R. Soc. Lond.* **1984,** *A393*, 171–183.

[19] W. R. Fright, R. H. T. Bates, J. M. Rowe, D. S. Spencer, M. G. Somekh, G. A. D. Briggs, Reconstruction of the complex reflectance function in acoustic microscopy. *J. Microsc.* **1989,** *153*, 103–117.

[20] R. N. Bracewell, *The Fourier Transform and Its Applications*. New York: McGraw-Hill **1978**.

[21] K. K. Liang, S. D. Bennett, G. S. Kino, Precision phase measurements with short tone burst signals in acoustic microscopy. *Rev. Sci. Instrum.* **1986,** *57*, 446–452.

[22] G. A. D. Briggs, M. G. Somekh, C. Ilett, Acoustic microscopy in materials science, in: *Microscopy—Techniques and Capabilities*, SPIE, Vol. 368 (Ed. L. R. Baker). Bellingham, WA: SPIE **1982,** pp. 74–80.

[23] Y. Tsukahara, E. Takeuchi, E. Hayashi, Y. Tani, A new method of measuring surface layer-thickness using dips in angular dependence of reflection coefficients. *IEEE 1984 Ultrasonics Symp. Proc.* Piscataway, NJ: IEEE **1984,** pp. 992–996.

[24] Y. Tsukahara, N. Nakaso, J. Kushibiki, N. Chubachi, An acoustic micrometer and its application to layer thickness measurements. *IEEE Trans. Ultrason. Ferroelect. Freq. Control* **1989**, *UFFC-36*, 326–331.

[25] M. G. Somekh, G. A. D. Briggs, C. Ilett, The effect of anisotropy on contrast in the scanning acoustic microscope. *Phil. Mag.* **1984,** *49*, 179–204.

[26] R. W. Gerchberg, W. O. Saxton, A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* **1972,** *35*, 237–246.

[27] R. H. T. Bates, M. J. McDonnell, *Image Restoration and Reconstruction*. Oxford: Clarendon Press **1986**.

[28] W. Parmon, H. L. Bertoni, Ray interpretation of the material signature in the acoustic microscope. *Electron Lett.* **1979,** *15*, 684–686.

[29] H. L. Bertoni, Ray-optical evaluation of *V(z)* in the reflection acoustic microscope. *IEEE Transonics Ultrason.* **1984,** *SU-31*, 105–116.

[30] G. A. D. Briggs, J. M. Rowe, A. M. Sinton, D. S. Spencer, Quantitative methods in acoustic microscopy. *IEEE 1988 Ultrasonics Symp. Proc.* Piscataway, NJ: IEEE **1988,** pp. 743–749.

[31] J. M. Rowe, Quantitative acoustic microscopy of surfaces. D. Phil. Thesis, Oxford University **1988**.

[32] J. M. Saurel, A. Saied, J. Attal, P. Dargent, La Microscopie acoustique: quelques aspects de son utilisation, in: *Méthodes avancées de charactérisation des surfaces* (Ed. D. David). Paris: Eyrolles **1991,** pp. 128–154.

[33] P. J. Burnett, G. A. D. Briggs, The elastic properties of ion-implanted silicon. *J. Mater. Sci.* **1986,** *21*, 1828–1836.

[34] J. Gomez Morales, R. Rodriguez, J. Durand, H. Ferdj-Allah, Z. Hadjoub, J. Attal, A. Doghmane, Characterization and identification of bernilite crystals by acoustic microscopy. *J. Mater. Res.* **1991,** *6*, 2484–2489.

[35] J. Kushibiki, A. Ohkubo, N. Chubachi, Linearly focused acoustic beams for acoustic microscopy. *Electron. Lett.* **1981,** *17*, 520–522.

[36] J. D. Achenbach, V. S. Ahn, J. G. Harris, Wave analysis of the acoustic material signature for the line focus microscope. *IEEE Trans. Ultrason. Ferroelect. Freq. Control UFFC-38* **1991,** 380.

[37] J. Kushibiki, N. Chubachi, Material characterization, by line-focus-beam acoustic microscope. *IEEE Trans. Sonics Ultrason.* **1985**, *SU-32*, 189–212.

[38] K. Dransfeld, E. Salzmann, Excitation, detection and attenuation of high-frequency elastic surface waves, in: *Physical Acoustics VII* (Eds. W. P. Mason, R. N. Thurston). New York: Academic **1970,** pp. 260–283.

[39] C. B. Scruby, K. R. Jones, L. Antoniazzi, Diffraction of elastic waves by defects in plates. *J. NDE* **1987,** *5*, 145–156.

[40] P. M. Burnett, G. A. D. Briggs, S. M. Al-Shukri, J. F. Duffey, R. M. De La Rue, Acoustic properties of proton-exchanged $LiNbO_3$ studied using the acoustic microscope $V(z)$ technique. *J. Appl. Phys.* **1986,** *60*, 2517–2522.

[41] C. K. Jen, The role of acoustic properties in designs of acoustic and optical fibres. *Mater. Sci. Eng.* **1989**, *A122*, 1–8.

[42] J. Kushibiki, Y. Matsumoto, M. Satake, N. Chubachi, Nondestructive evaluation of acoustic inhomogeneity on wagers by line-focus-beam acoustic microscope. *Ultrason. Int.* **1985,** *85*, 809–814.

[43] J. Kushibiki, H. Takahashi, T. Kobayashi, N. Chubachi, Characterization of $LiNbO_3$ crystals by line-focus-beam acoustic microscopy. *Appl. Phys. Lett.* **1991,** *58*, 2622–2624.

[44] J. Kushibiki, H. Takahashi, T. Kobayashi, N. Chubachi, Quantitative evaluation of elastic properties of $LiTaO_3$ crystals by line-focus-beam acoustic microscopy, *Appl. Phys. Lett.* **1991,** *58*, 893–895.

[45] J. Kushibiki, H. Asano, T. Ueda, N. Chubachi, Application of line-focus-beam acoustic microscope to inhomogeneity detection on SAW device materials. *IEEE 1986 Ultrasonics Symp. Proc.* Piscatoway, NJ: IEEE **1986,** pp. 749–753.

[46] R. B. Thompson, Y. Li, W. A. Spitzig, G. A. D. Briggs, A. F. Fagan, J. Kushibiki, Characterization of the texture of heavily deformed metal–metal composites with acoustic microscopy, in: *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 9 (Eds. D. O. Thompson, D. E. Chimenti). New York, London: Plenum **1990,** pp. 1433–1440.

[47] C. M. Sayers, Angular dependence of the Rayleigh wave velocity in polycrystalline metals with small anisotropy. *Proc. R. Soc. London* **1985,** *A400*, 175–182.

[48] D. A. Davids, D. Chizhik, H. L. Bertoni, Measured characteristics of an acoustic microscope having a bow-tie transducer. *IEEE 1988 Ultrasonics Symp. Proc.* Piscataway, NJ: IEEE **1988,** pp. 763–766.

[49] N. Chubachi, Ultrasonic microspectroscopy via Rayleigh waves, in: *Rayleigh-Wave Theory and Applications* (Eds. E. A. Ash, E. G. S. Paige). Berlin: Springer **1985,** pp. 291–297.

[50] J. Kushibiki, N. Chubachi, E. Tejima, Quantitative evaluation of materials by directional acoustic microscope. *Ultrason. Int.* **1989,** *89*, 736–743.

[51] N. Chubachi, H. Kanai, T. Sannomiya, T. Wakahara, Acoustic microscope for measuring acoustic properties by micro-defocusing method, in: *Acoustical Imaging*, Vol. 19 (Eds. H. Ermert, H.-P. Harjes). New York: Plenum **1992,** pp. 685–689.

[52] C.-H. Chou, B. T. Khuri-Yakub, Design and implementation of mixed-mode transducers, *IEEE Trans. Ultrason. Ferroelect. Freq. Control UFFC-36*, **1989,** 337–341.

[53] S. W. Meeks, D. Peter, D. Horne, K. Young, V. Novotny, Microscopic imaging of residual stress using a scanning phase-measuring acoustic microscope. *Appl. Phys. Lett* **1989,** *55*, 1835–1837. Cf. Residual stress mapping with a scanning phase-measuring microscope, *IEEE 1989 Ultrasonics Symp. Proc.* Piscataway, NJ: IEEE, pp. 809–812.

[54] G. W. Farnell, E. L. Adler, Elastic wave propagation in thin layers, in: *Physical Acoustics IX* (Eds. W. P. Mason, R. N. Thurston). New York: Academic **1972,** pp. 35–127.

[55] D. Walikainen, One GHz leaky SAW velocity of metal layers and bilayers evaporated onto fused quartz. *J. Vac. Sci. Technol.* **1992,** *A10* 290–294.

[56] J. Kushibiki, N. Chubachi, Application of LFB acoustic microscope to film thickness measurements. *Electron Lett.* **1987,** *23*, 652–654.

[57] J. Kushibiki, T. Ishikawa, N. Chubachi, Cut-off-characteristics of leaky Sezawa and pseudo-Sezawa wave modes for thin-film characterization. *Appl. Phys. Lett.* **1990,** *57*, 1967–1969.

[58] J. D. Achenbach, J. O. Kim, Y.-C. Lee, Measuring thin film elastic constants by line-focus acoustic microscopy in: *Advances in Microscopy 1* (Ed. G. A. D. Briggs). New York: Plenum **1995,** pp. 153–208.

[59] Z. Sklar, P. Mutti, N. C. Stoodley, G. A. D. Briggs, Measuring the elastic properties of stressed materials by quantitative acoustic microscopy, in: *Advances in Acoustic Microscopy 1* (Ed. G. A. D. Briggs). New York: Plenum **1995,** pp. 209–247.

[60] Y. Tsukahara, N. Nakaso, K. Ohira, M. Yanaka, Interaction of acoustic waves with solid surfaces, in: *Advances in Acoustic Microscopy 2* (Eds. G. A. D. Briggs, W. Arnold). New York: Plenum **1996**.

[61] A. Atalar, H. Köymen, A. Bozkurt, G. Yaraliŏglu, Lens geometries for quantitative acoustic microscopy, in: *Advances in Acoustic Microscopy 1* (Ed. G. A. D. Briggs). New York: Plenum **1995,** pp. 117–151.

[62] A. Atalar, L. Degertekin, H. Köymen, Acoustic parameter mapping of layered materials using a Lamb wave lens, in: *Acoustical Imaging*, Vol. 19 (Eds. H. Ermert, H.-P. Harjes). New York: Plenum **1992,** pp. 627–632.

[63] P. Mutti, C. E. Bottani, G. Ghislotti, M. Beghi, G. A. D. Briggs, J. R. Sandercock, Surface Brillouin spectroscopy—extending surface wave measurements to 20 GHz, in: *Advances in Acoustic Microscopy 1* (Ed. G. A. D. Briggs). New York: Plenum, **1995,** pp. 249–300.

[64] A. M. Sinton, G. A. D. Briggs, Y. Tsukahara, Time-resolved acoustic microscopy of polymer coatings, in: *Acoustical Imaging*. Vol. 17 (Eds. H. Shimizu, N. Chubachi, J. Kushibiki). New York: Plenum, **1989,** pp. 87–94.

[65] C. M. W. Daft, G. A. D. Briggs, Wideband acoustic microscopy of tissue. *IEEE Trans. Ultrason. Ferroelect. Freq. Control UFFC-36* **1989,** 258–263.

[66] G. A. D. Briggs, J. Wang, R. Gundle, Quantitative acoustic microscopy of individual living human cells. *J. Microsc.* **1993,** *172*, 3–12.

[67] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, *Numerical Recipes*. Cambridge: Cambridge University Press **1986**.

[68] G. S. Kino, *Acoustic waves: devices, imaging and analog signal processing*. Englewood Cliffs, NJ: Prentice-Hall **1987**.

[69] A. V. Oppenheim, R. W. Schafer, *Digital signal processing*. Englewood Cliffs, NJ: Prentice-Hall **1975**.

[70] D. Knauss, T. Zhai, G. A. D. Briggs, J. W. Martin, Measuring short cracks by time-resolved acoustic microscopy, in: *Advances in Acoustic Microscopy 1* (Ed. G. A. D. Briggs). New York: Plenum **1995,** pp. 49–77.

[71] M. G. Silk, Changes in ultrasonic defect location and sizing. *NDT Int.* **1987,** *20*, 9–14.

[72] J. P. Charlesworth, J. A. G. Temple, *Ultrasonic Time-of-Flight Diffraction*. New York: Wiley **1989**.

[73] J. M. R. Weaver, C. M. W. Daft, G. A. D. Briggs, A quantitative acoustic microscope with multiple detection modes. *IEEE Trans. Ultrason. Ferroelect. Freq. Control* **1989,** *36*, 554–560.

[74] D. Knauss, D. D. Bennink, T. Zhai, G. A. D. Briggs, J. W. Martin, Depth measurement of short cracks with an acoustic microscope. *J. Mater. Sci.*, in press **1992.**

[75] R. H. Tew, J. R. Ockendon, G. A. D. Briggs, Acoustical scattering by a shallow surface-breaking crack in an elastic solid under light fluid loading, in: *Recent Developments in Surface Acoustic Waves* (Eds. D. F. Parker, G. A. Maugin). Berlin: Springer **1988,** pp. 309–316.

[76] O. Kolosov, K. Yamanaka, Nonlinear detection of ultrasonic vibrations in an atomic force microscope. *Jpn. J. Appl. Phys.* **1993,** *32*, L1095–1098.

[77] O. Kolosov, H. Ogiso, H. Tokumoto, K. Yamanaka, Elastic imaging with nanoscale and atomic resolution, in: *Nanostructures and Quantum Effects* (Eds. H. Sakaki, H. Noge). Berlin: Springer-Verlag **1994,** *34*, pp. 345–8.

[78] K. Yamanaka, New approaches in acoustic microscopy for noncontact measurement and ultrahigh resolution, in: *Advances in Acoustic Microscopy 1* (Ed. G. A. D. Briggs). New York: Plenum **1995,** pp. 301–342.

[79] U. Rabe, W. Arnold, Acoustic microscopy by atomic force microscopy. *Appl. Phys. Lett.* **1994,** *64*, 1493–5.

[80] N. A. Burnham, A. J. Kulik, G. Gremaud, G. A. D. Briggs, Nanosubharmonics: the dynamics of small nonlinear contacts. *Phys. Rev. Lett.* **1995,** *74*, 5092–5095.

Part IV

Electron Microscopy

1 Stationary Beam Methods

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

1.1 Transmission Electron Microscopy

1.1.1 Diffraction Contrast Transmission Electron Microscopy

1.1.1.1 Introduction

Image formation in transmission electron microscopy is essentially a diffraction phenomenon, normal absorption only playing a minor role. It is therefore necessary to discuss first the basis of electron diffraction. A detailed interpretation of an image requires a knowledge of the corresponding diffraction pattern, adequately oriented with respect to the image. Modern transmission electron microscopes are constructed in such a way that they can be easily switched from the imaging mode to the diffraction mode, and vice versa, without changing the orientation of the specimen. The electron microscope optics are discussed briefly below, emphasizing the general principle rather than the concrete details, since the latter depend on the particular instrument.

1.1.1.2 Instrumentation

Transmission Microscopes

A modern transmission electron microscope can be schematized as a three-lens system: an objective lens, an intermediate lens and a projector lens. Each of these lenses is in actual fact often a composite lens, but since the details of this depend on the particular instrument this will not be discussed here. Moreover, condensor lenses are used in the illumination system, but we will consider only the ray paths in a three-lens, image formation system. The system allows easy switching from the high magnification imaging mode to the selected area diffraction mode. The ray paths are shown in Fig. 1. Movable selection apertures are placed: one in the image plane of the objective lens and a second one close to the back focal plane. The front aperture is used to select a small area ($\leqslant 1\ \mu m$) of the specimen whilst viewing the image. The second one enables us to select either a single beam or a number of image-forming diffracted beams. The image resolution of the system is to a large extent determined by the characteristics of the objective lens, in particular by its spherical aberration constant. Whereas in the high-resolution structure imaging mode the quality of the objective lens is crucial, this is much less the case for diffraction contrast images. In the latter mode the availability of very small beam-selection apertures is important, since this determines to what extent a single beam can be selected for dark-field imaging of materials with a relatively large unit cell.

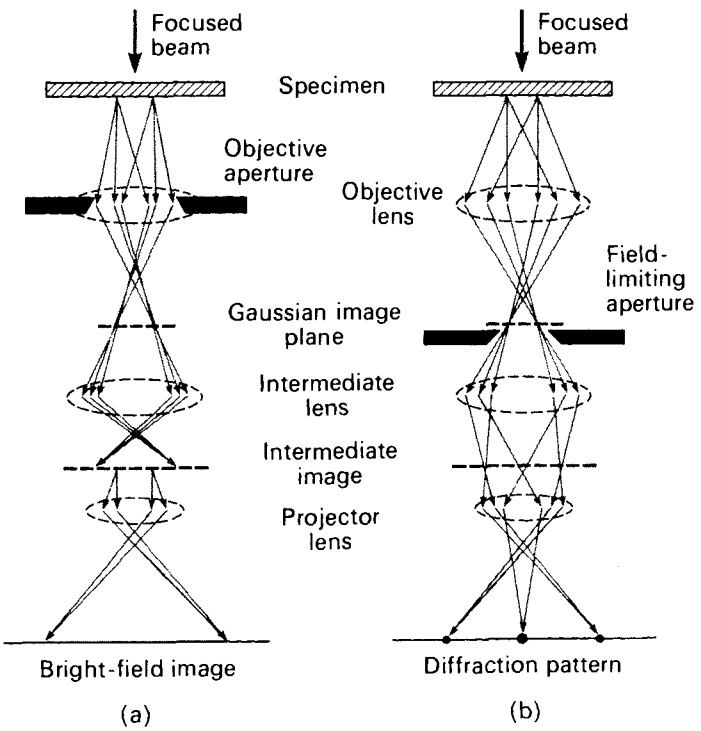


Figure 1. Ray paths in a reference transmission electron microscope. (a) High resolution high magnification imaging mode. (b) Selected area diffraction mode.

The intermediate and projector lenses provide the desired magnification. When using the high-resolution mode a sufficiently large magnification (on the fluorescent screen or on the TV monitor) is necessary to see separate atom columns so as to allow proper focusing. Although magnetic lenses normally rotate the image about the optical axis, in recently designed microscopes these rotations are compensated for by a suitable device and, as a result, the image and diffraction pattern have parallel orientations, which is particularly important in diffraction contrast work. In certain instruments the image and the diffraction pattern may still differ in orientation by 180° for certain lens combinations.

Lens Configurations

High Resolution, High Magnification Imaging Mode

The electron beam produced by an electron source (see below) is collimated by the condensor lens system (not shown in Fig. 1a) and scattered by the specimen. An image is formed in the image plane of the objective lens (Fig. 1a). The selector aperture allows us to select one area of the image (i.e., of the specimen) which is then magnified by the intermediate lens. The intermediate lens is focused on the image plane of the objective lens and an intermediate image is formed in the image plane of the intermediate lens. This image is the object for the projector lens which

forms a final image on a fluorescent screen or on the entrance plane of a recording device (see below).

Diffraction Mode

In the diffraction mode (Fig. 1b) the intermediate lens is weakened, that is the focal length is made larger, in such a way that the back focal plane of the objective lens coincides with the object plane of the projector lens. A magnified representation of the diffraction pattern is then produced on the fluorescent screen. In the process the selected area is not changed since only the strength of the intermediate lens has been modified. The diffraction pattern is thus representative of the selected area. However, it should be noted that under high resolution conditions the field of view in the image is much smaller than the selected area in the diffraction mode.

Diffraction Contrast Imaging Modes

As discussed below, diffraction contrast images are maps of the intensity distribution in highly magnified diffraction spots. They are usually obtained under two-beam conditions. The aperture placed close to the back focal plane of the objective lens allows us to select either the transmitted beam or the diffracted beam. The corresponding diffraction spot is subsequently magnified by the rest of the lens system. If the transmitted beam is selected, a bright field image is obtained; that is the area of the image not covered by the specimen is bright. If the diffracted beam is selected, a dark field image is obtained; the background is now dark.

Whereas the beam remains along the optical axis in the case of a bright field image, it encloses twice the Bragg angle of the active reflection with the optical axis for a dark field image. Non-axial beams suffer angle-dependent lens aberrations and the corresponding image is therefore often blurred by streaking. This can be avoided by tilting the incident beam over twice the Bragg angle; the diffracted beam then travels along the optical axis. Recently developed microscopes have a built-in device that allows the incident beam to be deflected over the required angle to bring a selected diffracted beam along the optical axis.

Electron Sources

In older microscopes the source of electrons is a heated V-shaped tungsten wire, the tip of the V forming a localized emitter of electrons with an effective size of the order of 10 μm. In subsequent versions the effective size of the emitting area is decreased by attaching a pointed tungsten wire to the tip of the V-shaped heating filament. Also, a pointed LaB_6 single crystal mounted on a V-shaped tungsten heating filament is often used because of its small work function. It emits electrons at a lower temperature than tungsten and, moreover, the thermal energy distribution of the electrons is narrower ($\approx$1 eV), thus leading to less chromatic abberation. In recent years the thermal spread of the emitted electrons has been reduced further to less than 0.5 eV by the use of cold field emission guns. Such guns consist of a very fine point placed on the pointed filament which emits electrons by tunneling. A sufficiently strong electric field (of the order of 10^6 V cm^{-1}) is required in the vicinity of the field point to cause a sufficiently strong

tunneling current density. Tunneling can be thermally assisted by slightly heating the tungsten point. The brightness of such sources is about a thousand times larger than that of the reference tungsten filament but, since the emitting area ($\sim 5\,mm^2$) is much smaller, the electron current is smaller. Field emission guns are particularly important in applications where a high degree of coherence is desirable, such as in high resolution atomic imaging.

Electron Beam Shaping

The electrons emitted by the source must be accelerated and a collimated beam must be formed. This is achieved by applying a large negative voltage to the emitter, keeping the anode grounded. Close to the emitter is a cylindrical electrode, the 'Wehnelt cylinder', which is biased at a negative voltage of a few hundred volts with respect to the emitter. The function of the Wehnelt cylinder is to stabilize the beam current and to focus the electron beam so as to form behind the anode a cross-over, which acts as the virtual source of electrons. The incident beam is finally shaped into a parallel beam (or possibly a convergent beam) by the system of condensor lenses and by apertures. Typically, the angular spread may be made as small as 10^{-4} rad.

Electrical Supply System

Very stable supply systems for the high voltage used to accelerate the electrons and for the lens currents are essential to minimize chromatic abberations. In present day commercial microscopes, supply systems with a stability of 1 part in 10^6 are used.

Vacuum Systems

Clean and vibration-free vacuum systems are essential to provide stability and avoid contamination of the specimen by a carbon film resulting from the cracking of organic molecules present in the residual gas. Anti-contamination devices such as metal blades surrounding the specimen and cooling to liquid nitrogen temperature are available in most microscopes. Use is made of diffusion pumps, turbomolecular pumps, ion pumps, and even sublimation pumps to achieve a vacuum of the order of 133×10^{-8} Pa (10^{-8} torr).

Recording and Viewing Media

Usually images are made visible on a fluorescent screen and viewed by means of binoculars. The simplest recording medium is still the photographic emulsion which is particularly efficient for recording electron images. A shutter mechanism allows one to remove the fluorescent screen and expose the underlying film. Degassing the photographic material prior to use is strongly recommended. In order to reduce radiation damage to the specimen it is important to keep the electron dose used to view and record the image as small as possible; for this purpose electronic viewing and recording methods have been developed. These techniques are discussed in Chap. 8, Sec. 1 and 2 of this Volume.

1.1.1.3 Electron Diffraction

Atomic Scattering Factor

Electrons are scattered by atoms as a result of the Coulomb interaction with the positively charged nucleus and with the negatively charged electron cloud. The atomic scattering factor thus contains two contributions of opposite sign [1]

$$f_e(\theta) = \frac{me^2\lambda}{2h^2}\,\frac{[Z - f_x(\theta)]}{\sin^2\theta} \qquad (1)$$

where $f_x(\theta)$ is the atomic scattering factor for X-rays, which are only scattered by the electron cloud. Z is the atomic number (or the positive charge on the nucleus), λ is the electron wavelength [see Eq. (2)], m is the electron mass, e is the electron charge, h is Planck's constant, and θ is the scattering angle. The electron wavelength is given by the de Broglie relation $\lambda = h/mv$. It is related to the accelerating potential E by the relativistic relation

$$\lambda = h\left[2m_0E\left(1 + \frac{eE}{2m_0c^2}\right)\right]^{-1/2} \qquad (2)$$

where m_0 is the rest mass of the electron. With E in the range 200–400 kV the electrons used in electron microscopy are to a non-negligible extent relativistic since they travel at speeds larger than half the speed of light.

The first term in Eq. (1) clearly relates to the nucleus, whereas the second term is due to the electron cloud. The interaction with matter is much stronger for electrons than for X-rays or neutrons by a factor of about 10^4. Multiple diffraction events will therefore have a high probability. The factor $1/\sin^2\theta$ in Eq. (1) causes scattering to be oriented mainly in the forward direction. Values of $f_c(\theta)$ for different atoms have been tabulated in [2].

Diffraction by Crystals

The amplitude diffracted in any given direction by an assembly of atoms is the sum of the amplitudes scattered by the individual atoms in that direction, taking into account the phase differences resulting from the geometry of the assembly. In a crystal the atoms are located on a three-dimensional lattice which can be characterized by its three base vectors $\boldsymbol{a}_1, \boldsymbol{a}_2$, and $\boldsymbol{a}_3$. A general lattice node is then given by

$$\boldsymbol{A}_L = l_1\boldsymbol{a}_1 + l_2\boldsymbol{a}_2 + l_3\boldsymbol{a}_3 \qquad (3)$$

where l_j are integers. The volume of the unit cell is $V_a = (\boldsymbol{a}_1 \times \boldsymbol{a}_2)\cdot\boldsymbol{a}_3$. The sum and difference of two lattice vectors is again a lattice vector.

It is convenient to define the corresponding reciprocal lattice by its base vectors $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, and $\boldsymbol{b}_3$, which are related to the $\boldsymbol{a}_j$ by the relations $\boldsymbol{a}_i \cdot \boldsymbol{b}_j = \delta_{ij}$ where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. A general node point of the reciprocal lattice is then given by

$$\boldsymbol{B}_H = h_1\boldsymbol{b}_1 + h_2\boldsymbol{b}_2 + h_3\boldsymbol{b}_3 \qquad (4)$$

the h_j are integers, called 'Miller indices' for planes.

A useful property is $\boldsymbol{A}_L \cdot \boldsymbol{B}_H = \text{Integer}$. The volume of the unit cell of the reciprocal lattice is $V_b = 1/V_a$. The spacing of lattice planes with indices $\boldsymbol{H}$ (h_1, h_2, h_3) is given by $d_H = 1/|\boldsymbol{B}_H|$.

A crystal structure is described with respect to its lattice by specifying the contents of the unit cell, that is by giving the position vectors $\boldsymbol{\rho}_j$ for all atoms $j = 1, \ldots, N$ in the unit cell. The assembly

of scattering units is thus given by the position vectors $\boldsymbol{A}_L + \boldsymbol{\rho}_j$ for the atoms of type j.

Diffraction Conditions

The diffraction conditions for a periodic object can be formulated in terms of direct space (i.e., of the lattice) or in terms of diffraction space (i.e., of the reciprocal lattice). The two formulations have the same physical content but the latter is often more convenient and more directly related to the diffraction pattern.

If the attention is focused on the lattice the diffraction condition is Bragg's law [3] which states that the path difference between waves 'reflected' by successive lattice planes is an integer number of wavelengths

$$2d_H \sin\theta_H = n\lambda \tag{5a}$$

where d_H is the interplanar spacing and θ_H is the corresponding Bragg angle (Fig. 2); n is an integer. This statement might create the wrong impression that 'reflection' takes place. However, the difference with specular reflection is important: only for the discrete angles θ_H does 'reflection' take place, whereas in specular reflection all angles are permitted. This clearly shows that we have, in fact, interference rather than reflection. Since the notion of 'reflection' has nevertheless been extremely useful, especially in structure determination, it continues to be used. Due to the small wavelength of electrons [$\approx 3 \times 10^{-3}$ nm (2×10^{-2} Å)] Bragg angles are quite small ($\approx 10^{-3}$ rad) and Bragg's law can often be approximated by

$$2d_H\theta_H = n\lambda \tag{5b}$$

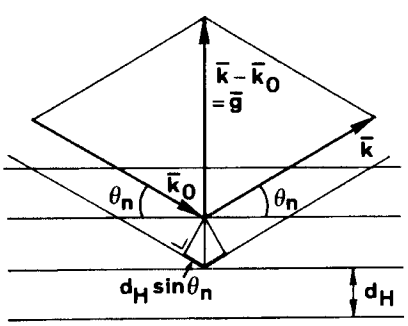


Figure 2. Geometry of Bragg scattering ($\theta_n = \theta_H$).

In reciprocal space the diffraction conditions can be formulated in terms of Ewald's sphere

$$\boldsymbol{k}_g = \boldsymbol{k}_0 + \boldsymbol{g} \tag{6}$$

where $\boldsymbol{k}_0$ is the wavevector of the incident plane wave [$\boldsymbol{k}_0 = (1/\lambda)\boldsymbol{e}_n$, where $\boldsymbol{e}_n$ is the normal to the plane wavefront] and $\boldsymbol{k}_g$ is the wavevector of the scattered wave. The term $\boldsymbol{g}$ is a reciprocal lattice vector, called the *diffraction vector*; it is an element of the set $\boldsymbol{B}_H$ and is thus specified by three integers h_1, h_2, and h_3, the Miller indices.

The Ewald condition gives rise to an elegant construction (Fig. 3). Let $\boldsymbol{k}_0$ represent the incident wave; its endpoint coincides with the origin of reciprocal space O: its starting point C is then the center of a sphere with radius $1/\lambda$ (Ewald's sphere) [4]. If this sphere passes through another reciprocal lattice node H, a diffracted beam $\mathrm{CH} \equiv \boldsymbol{k}_g$ is produced. Tilting the specimen is equivalent to tilting the reciprocal lattice over the same angle about a parallel axis. Tilting thus permits the 'excitation' of specific nodes of the reciprocal lattice. It should be noted that since the electron wavelength is of the order of 0.001 nm (0.01 Å) (i.e., $|\boldsymbol{k}| = 10\,\mathrm{nm}^{-1}$ ($10^2\,\text{Å}^{-1}$), whereas the mesh size of the reciprocal lattice is of

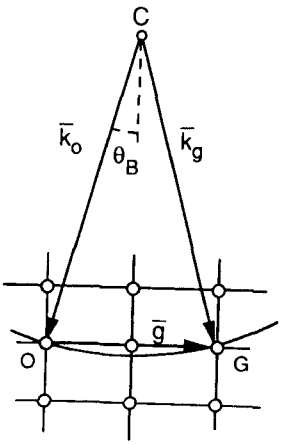


Figure 3. Ewald construction. The Ewald sphere with radius $|\boldsymbol{k}_0| = |\boldsymbol{k}_g| = 1/\lambda$ passes through the node G. $\boldsymbol{k}_0$: wavevector incident beam; $\boldsymbol{k}_g$: wavevector diffracted beam; $\boldsymbol{g}$: diffraction vector.

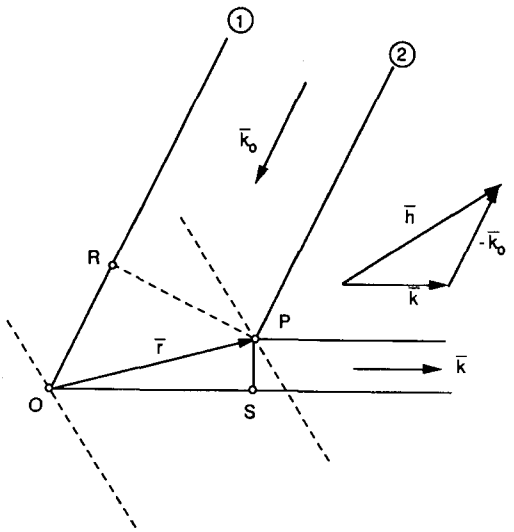


Figure 4. Path difference OR + OS between waves scattered at O and at P.

the order of 0.1 nm (1 Å^{-1}), the sphere radius is quite large and it can be approximated for most practical purposes by a plane normal to the incident wavevector $\boldsymbol{k}_0$. The diffraction pattern is thus obtained as a central projection of a planar section of reciprocal space on to the screen (or photographic plate).

Diffraction Amplitude

Let $\boldsymbol{k}_0$ be the wavevector of an incident wave $\exp(2\pi i \boldsymbol{k}_0 \boldsymbol{r})$ and $\boldsymbol{k} = 1/\lambda$ the wavevector of the scattered wave. The phase difference between the waves diffracted by an atom at the origin O and an atom P at $\boldsymbol{r}_j$ is $(2\pi/\lambda)(\mathrm{OR} + \mathrm{OS}) = 2\pi(\boldsymbol{k} - \boldsymbol{k}_0) \cdot \boldsymbol{r}_j$ (Fig. 4) and the total scattered amplitude by the assembly of points at positions $\boldsymbol{r}_j$ is then given by

$$A(\boldsymbol{k}) = \sum_j f_j \exp[2\pi i(\boldsymbol{k} - \boldsymbol{k}_0) \cdot \boldsymbol{r}_j] \qquad (7)$$

where f_j is the atomic scattering amplitude of the atoms at $\boldsymbol{r}_j$. This amplitude will be maximal if all waves are in phase. If the atoms are assumed to be located on a lattice $\boldsymbol{r}_j = \boldsymbol{A}_L$; a maximum of $A(\boldsymbol{k})$ will then occur if all exponents are integer multiples of $2\pi i$, that is, if $(\boldsymbol{k} - \boldsymbol{k}_0) \cdot \boldsymbol{A}_L =$ Integer. This will be the case if $\boldsymbol{g} \equiv \boldsymbol{k} - \boldsymbol{k}_0 = \boldsymbol{B}_H$, which is the Ewald condition [Eq. (6)].

The scattered amplitude will also differ from zero even if the *scattering vector* $\boldsymbol{g} \equiv \boldsymbol{k} - \boldsymbol{k}_0$ differs somewhat from a reciprocal lattice vector. This deviation from the exact diffraction condition is described by the vector $\boldsymbol{s}$ (Fig. 5) ('excitation error' also called 'deviation parameter'). It measures the distance of the reciprocal lattice node $\boldsymbol{G}_H$ to the Ewald sphere; it is normal to the foil surface. The vector $\boldsymbol{s}$ is positive when the reciprocal lattice node is inside Ewald's sphere; it is negative when it is outside. The positive direction of $\boldsymbol{s}$ is thus in the direction of propagation of the electrons. Equation (7) can now be

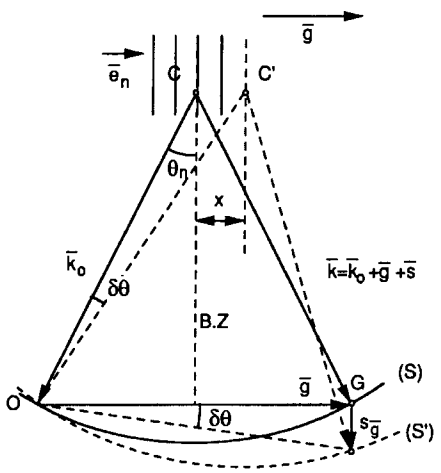


Figure 5. The deviation from the exact Bragg condition is characterized by s_g, which is positive in the sense of the propagating electrons; x is an alternative deviation parameter. BZ: Brillouin zone boundary.

generalized to

$$A(\boldsymbol{g}) = \sum_j f_j \exp[2\pi i(\boldsymbol{g}+\boldsymbol{s})\cdot\boldsymbol{r}_j] \tag{8}$$

Taking into account that an atom with a scattering factor f_j is found at the positions $\boldsymbol{A}_L + \boldsymbol{\rho}_j$ $(j = 1, \ldots, N)$, the amplitude can be written as

$$\begin{aligned} A(\boldsymbol{g}) &= \sum_L \sum_{j=1,\ldots,N} f_j \\ &\quad \times \exp[2\pi i(\boldsymbol{g}+\boldsymbol{s})\cdot(\boldsymbol{A}_L+\boldsymbol{\rho}_j)] \\ &= \sum_L \left[\sum_j f_j \exp(2\pi i \boldsymbol{g}\cdot\boldsymbol{\rho}_j)\right] \\ &\quad \times \exp(2\pi i \boldsymbol{s}\cdot\boldsymbol{A}_L) \end{aligned} \tag{9}$$

where we note that $\boldsymbol{g}\cdot\boldsymbol{A}_L = \text{Integer}$, since $\boldsymbol{g}$ is an element of the set $\boldsymbol{B}_H$ and $\boldsymbol{s}\cdot\boldsymbol{\rho}_j$ is negligible compared to the other terms.

The expression

$$F_g \equiv \sum_j f_j \exp(2\pi i \boldsymbol{g}\cdot\boldsymbol{\rho}_j)$$

describes the amplitude scattered by one unit cell in the direction defined by $\boldsymbol{g}$; it is called the *structure amplitude* or *structure factor*. Equation (9) then becomes

$$A(\boldsymbol{g}) = F_g \sum_L \exp(2\pi i \boldsymbol{s}\cdot\boldsymbol{A}_L) \tag{10}$$

We assume that the crystal contains N_1, N_2 and N_3 unit cells respectively along the $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$ directions. The scattering amplitude then becomes

$$\begin{aligned} A(\boldsymbol{g}) &= F_g \sum_{l_1=0}^{N_1-1} \sum_{l_2=0}^{N_2-1} \sum_{l_3=0}^{N_3-1} \\ &\quad \times \exp[2\pi i(s_1 l_1 a_1 + s_2 l_2 a_2 + s_3 l_3 a_3)] \end{aligned} \tag{11}$$

where s_1, s_2 and s_3 are the components of $\boldsymbol{s}$ along the three base directions of the lattice and $\boldsymbol{s}\cdot\boldsymbol{A}_L = s_1 l_1 a_1 + s_2 l_2 a_2 + s_3 l_3 a_3$. Separating the triple sum into the product of three single sums, one obtains

$$\begin{aligned} A(\boldsymbol{g}) = F_g \Bigg[&\sum_{l_1=0}^{N_1-1} \exp(2\pi i s_1 a_1 l_1) \\ &\cdot \sum_{l_2=0}^{N_2-1} \exp(2\pi i s_2 a_2 l_2) \\ &\cdot \sum_{l_3=0}^{N_3-1} \exp(2\pi i s_3 a_3 l_3) \Bigg] \end{aligned} \tag{12}$$

Performing the summation of the geometrical series, and omitting an irrelevant phase factor, one obtains finally the von Laue interference function

$$\begin{aligned} A(\boldsymbol{g}) \approx F_g &\frac{\sin(\pi s_1 N_1 a_1)}{\sin(\pi s_1 a_1)} \cdot \frac{\sin(\pi s_2 N_2 a_2)}{\sin(\pi s_2 a_2)} \\ &\cdot \frac{\sin(\pi s_3 N_3 a_3)}{\sin(\pi s_3 a_3)} N_1 N_2 N_3 \end{aligned} \tag{13}$$

which describes the dependence of the scattered amplitude on the deviation parameter.

Since $\pi s_1 a_1, \pi s_2 a_2$, and $\pi s_3 a_3$ are small, the sine terms in the denominators can be replaced by their arguments. We further note that for large N, $\sin(\pi Nsa/\pi Nsa) \simeq \delta(s)$ where $\delta(s) = 0$ for $s \neq 0$ and $\delta(s) = 1$ if $s = 0$. Taking this into account we can write

$$A(\boldsymbol{g}) = F_{\boldsymbol{g}}\, \delta(s_1)\, \delta(s_2)\, \delta(s_3)\, \frac{\Omega}{V_a} \tag{14}$$

where Ω is the volume of the crystal and V_a is the volume of the unit cell: $\Omega = N_1 N_2 N_3 V_a$.

With regard to transmission electron microscopy, thin foil specimens contain a large number of unit cells in the two lateral directions, but along the thickness the number of unit cells is finite and Eq. (13) reduces to

$$A(\boldsymbol{g}) = F_{\boldsymbol{g}} \frac{\Omega}{V_a}\, \delta(s_1)\, \delta(s_2)\, \frac{\sin(\pi N_3 a_3 s_3)}{(\pi N_3 a_3 s_3)} \tag{15}$$

where s_3 is measured perpendicular to the foil. Introducing the specimen thickness $t = N_3 a_3$ we can write

$$A(\boldsymbol{g}) = F_{\boldsymbol{g}} \frac{S_a}{V_a}\, \frac{\sin(\pi s_3 t)}{\pi s_3}\, \delta(s_1)\, \delta(s_2) \tag{16a}$$

with $S_a = N_1 N_2 a_1 a_2$. This can be rewritten per unit surface area as:

$$A(\boldsymbol{g}) = \frac{\sin(\pi s t)}{s t_g} \tag{16b}$$

with $t_g = (\pi V_a / F_{\boldsymbol{g}})$; t_g is called the *extinction distance*.

This result is interpreted in terms of diffraction space as meaning that the reciprocal lattice nodes of a thin foil are not mathematical points but are rod shaped (relrods) and perpendicular to the foil plane with a weight given by $(\sin \pi s t)/\pi s$. This function is shown in Fig. 6.

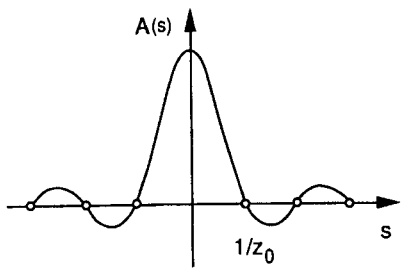


Figure 6. Scattered amplitude versus s according to the kinematical approximation. z_0: foil thickness.

The corresponding intensity profile is called the 'rocking curve' according to the kinematical theory (Fig. 6). An amplitude can be associated with each intersection point of the Ewald sphere with this segment, the amplitude being given by the value of this profile at the intersection point. It is customary to describe the diffraction geometry by saying that the reciprocal lattice nodes have become 'relrods' with a length profile in the z direction, parallel to the foil normal, given by the above-mentioned function, but infinitely sharp in the directions x and y in the foil. In a thin foil the vector s is thus oriented along the normal to the foil plane. By convention, s is counted positive in the sense of the propagating electrons, that is if the reciprocal lattice node G is inside Ewald's sphere.

The Column Approximation

Since Bragg angles are quite small ($\approx 10^{-3}$ rad) in electron diffraction, the intensity observed at a point in the exit face of a foil is essentially determined by the diffraction events occurring in a narrow column of crystal centered on the considered point extending through the thickness of the foil in a direction parallel

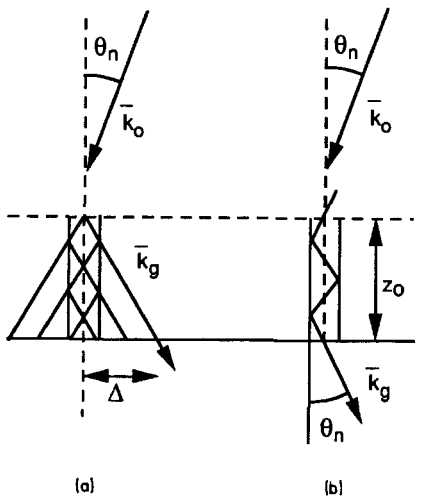


Figure 7. Column approximation: (a) kinematical approximation; (b) dynamical approximation.

to the incident beam (Fig. 7) [5]. Such a column diffracts independently of the surrounding columns; its lateral dimensions are, at most, given by $\Delta = \theta_n z_0$ ($\theta_n \equiv \theta_B$; $z_0 = t =$ thickness). For a strain-free foil limited by parallel faces, all such columns behave in the same manner and a uniform intensity results in the image. However, if defects are present, columns in the vicinity of the defects will produce an intensity which differs from that of the surrounding columns and hence a defect image will result. The amplitude found at the exit point of a column is given by (omitting irrelevant factors)

$$A_g = F_g \int_0^t e^{2\pi i s z}\, dz \tag{17}$$

where summation over all unit cells in the column has been replaced by integration along the columns. The meaning of z and t is shown in Fig. 8. The exponential is the phase factor, relative to the entrance point due to the volume element dz at level z behind the entrance face. If s is a constant,

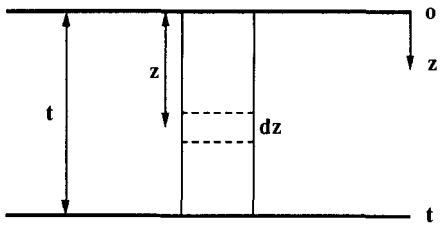


Figure 8. Illustration of the notations used.

as is the case in a defect-free crystal, the result is

$$A_g = \frac{F_g \sin(\pi s t)}{\pi s} \tag{18}$$

which is consistent with Eq. (16). However, if a defect is present, s becomes a function of x, y, and z. For a column at (x, y) one obtains

$$A_g(x, y) = F_g \int_0^t e^{2\pi i s(x,y,z)}\, dz \tag{19}$$

Each column, of which the size is arbitrarily small, now furnishes a picture element of the defect image. For most defects such images will have to be computed numerically since the analytical expressions may become too complicated. Defect imaging is discussed in detail below.

If the strain-free foil has a nonuniform thickness the intensity at the exit face will depend on the length of the column and it will be equal for columns of equal length. The columns that produce extrema in intensity form geometrical loci, called *thickness extinction contours*, which are dark for a minimum and bright for a maximum. In a wedge-shaped crystal such contours form a set of straight fringes parallel to the wedge edge, described by Eq. (18). Their formation is represented schematically in Fig. 9.

When a defect-free foil of uniform thickness is bent the s value becomes

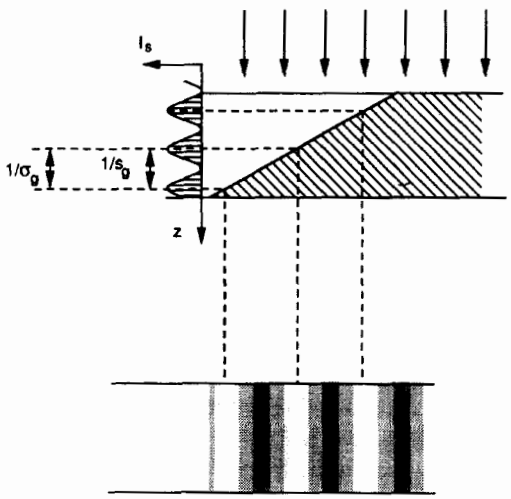


Figure 9. Model for the formation of thickness extinction contours at a wedge-shaped crystal. I_s: scattered intensity; s_g: excitation error; $\sigma_g = \sqrt{1+(s_g t_g)^2}/t_g$.

variable along a line perpendicular to the bending axis. The loci of equal s value are then again imaged as contours of equal brightness; they are called *inclination extinction contours* (Fig. 10a). These image in a sense the rocking curve represented by Eq. (18). Figure 10b shows a cylindrically bent uniformly thick foil of graphite; the inclination contours corresponding to the main and subsidiary maxima can clearly be observed.

Amplitude–Phase Diagram

A plane wave represented by

$$A \exp \mathrm{i}(kx - \omega t + \psi)$$

$$\equiv A \exp(\mathrm{i}\psi) \exp[\mathrm{i}(kx - \omega t)]$$

is characterized by a complex amplitude $A\exp(\mathrm{i}\psi)$, a wavevector $\boldsymbol{k}$ and an angular frequency ω. The interference between two waves of this type, assuming the wavevector $\boldsymbol{k}$ and the angular frequency ω to

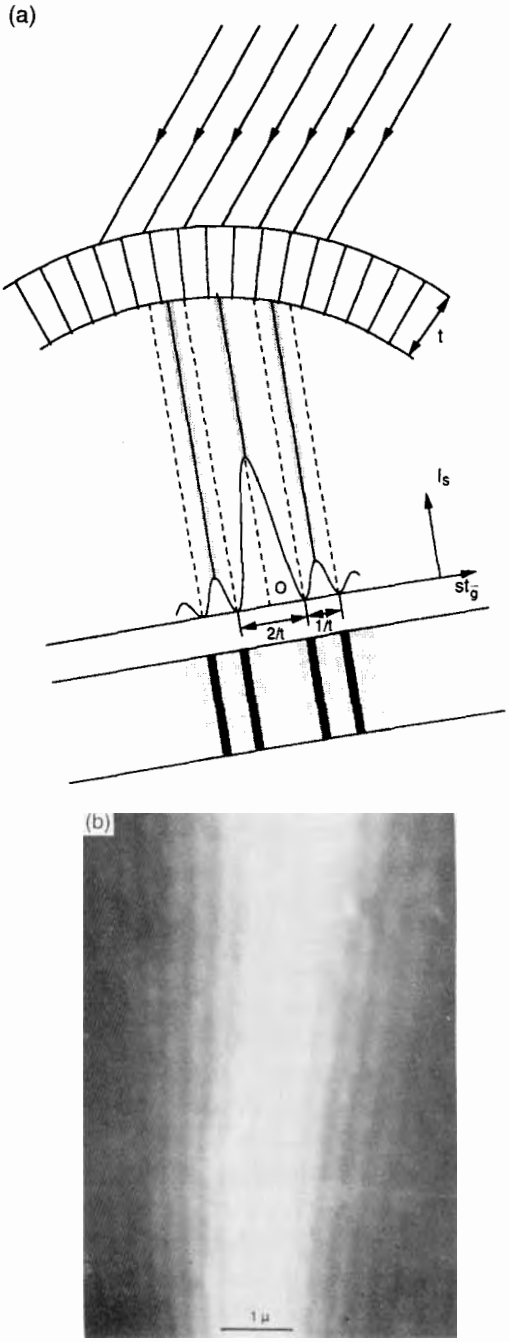


Figure 10. Diffracted intensity as a function of the angle of incidence in a cylindrically bent foil: (a) formation of bent contours (I_s: scattered intensity); (b) cylindrically bent graphite foil.

be the same for both, produces a resultant wave with the same ω and $\boldsymbol{k}$ but different A and ψ. The problem thus reduces to summing the complex amplitudes $A\exp(i\psi)$, the propagation factor being common to all waves.

The complex amplitude is represented in the complex plane by a vector with modulus A and argument ψ. It is easy to show that the sum of the two such vectors is again a vector representing the resultant wave. Waves with the same $\boldsymbol{k}$ and ω can thus be summed graphically by adding vectors in the complex plane.

Apart from a phase factor, the amplitude scattered by a column of crystal along the z axis, in the kinematical approximation, is given by the sum

$$A(\boldsymbol{g}) = F_g \sum_n \exp(2\pi i s z_n)\,\Delta z_n \qquad (20)$$

or in the continuum approximation by the integral mentioned above:

$$A(\boldsymbol{g}) = F_g \int_0^t \exp(2\pi i s z)\,\mathrm{d}z \qquad (21)$$

This sum can be considered as consisting of terms $F_g\,\Delta z \cdot \exp(2\pi i s z)$, that is $A = F_g\,\Delta z$; $\psi = 2\pi i s z$, corresponding with slices Δz of the column. The amplitude phase diagram then consists of vectors, all of the same length $F_g\,\Delta z$ and enclosing angles of $2\pi s\,\Delta z$. In the limit $\Delta z \to 0$, the locus of the endpoints is a circle with radius (Fig. 11)

$$R = F_g \lim_{\Delta z \to 0} (\Delta z / 2\pi s\,\Delta z)$$

$$= \frac{F_g}{2\pi s} \qquad (22)$$

The length of the circular arc is equal to the column length, that is to the foil thickness t. Figure 11 shows that: the diffracted

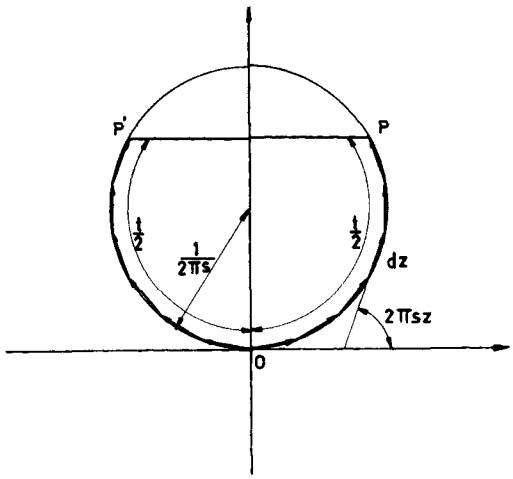


Figure 11. Amplitude phase diagram for a perfect crystal foil.

amplitude will be zero if the circular arc is a number of complete circles, that is for $t = k/s$; and there will be maxima if $t = (1/s)(k + \frac{1}{2})(k = \text{Integer})$, the maximum amplitude being equal to the diameter of the circle [i.e., $A_{\max} = (1/\pi s)F_h$].

Kikuchi Lines

In sufficiently thick and almost perfect specimens, spot patterns are no longer observed; instead a diffraction phenomenon, first discovered by Kikuchi in 1928 [6], is produced. It usually consists in the occurrence of pairs of bright and dark straight lines in the diffraction pattern, as shown in Fig. 12. In foils of intermediate thickness one can observe the Kikuchi pattern superimposed on the spot pattern. The geometry of the Kikuchi pattern can satisfactorily be explained by assuming that not only are electrons Bragg scattered, but that also a substantial fraction, especially in thick foils, is scattered inelastically and incoherently in the crystal, the energy

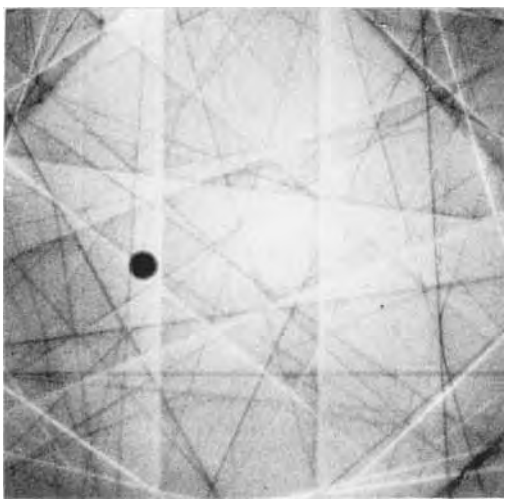


Figure 12. Pattern of Kikuchi lines in a rather thick silicon crystal.

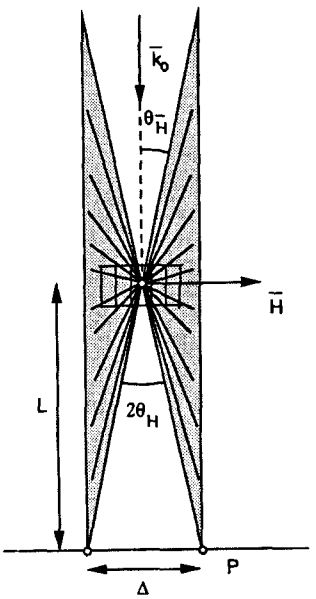


Figure 13. Geometry of the Kikuchi cones in the symmetrical orientation.

loss being much smaller than the energy of the incident electron; the electron wavelength is then not appreciably changed. Inside the crystal these randomly scattered electrons impinge on the lattice planes from all directions, but preferentially in the forward direction, and can subsequently give rise to Bragg scattering.

A symmetrical situation with respect to the set of lattice planes $\boldsymbol{H}$, with spacing d_H, is shown in Fig. 13. Bragg scattering out of the incident beam is assumed to be weak since the Bragg condition is not satisfied. However, a fraction of the randomly scattered electrons have the correct direction of incidence to give rise to Bragg diffraction by the set of lattice planes considered. The geometrical locus of these Bragg scattered electron beams is a double cone of revolution with an opening angle $(\pi/2) - \theta_H$ and with its axis along $\boldsymbol{H}$ (where θ_H is the Bragg angle). These cones are therefore rather 'flat' and the intersection lines of the two sheets of this double cone with the photographic plate P looks like two parallel straight lines, although in actual fact they are two branches of a hyperbolic conical section. The angular separation of these two lines is $2\theta_H$. The separation Δ observed on the plate is thus $\Delta = 2L\theta_H$, where L is the camera length, i.e. the specimen to plate distance. The angular separation does not depend on the crystal orientation.

The geometry of this cone (i.e., the axis of revolution and the opening angle) is entirely fixed by the crystal lattice and the electron wavelength, and is independent of the incident beam direction. Tilting the specimen thus leads to an equal tilt of the double cone, but leaves the geometry of the spot diffraction pattern unchanged, provided the same reflections remain excited, that is as long as the same 'relrods' are intersected by Ewald's sphere. The relative position of the spot pattern and of the Kikuchi line pattern is thus very

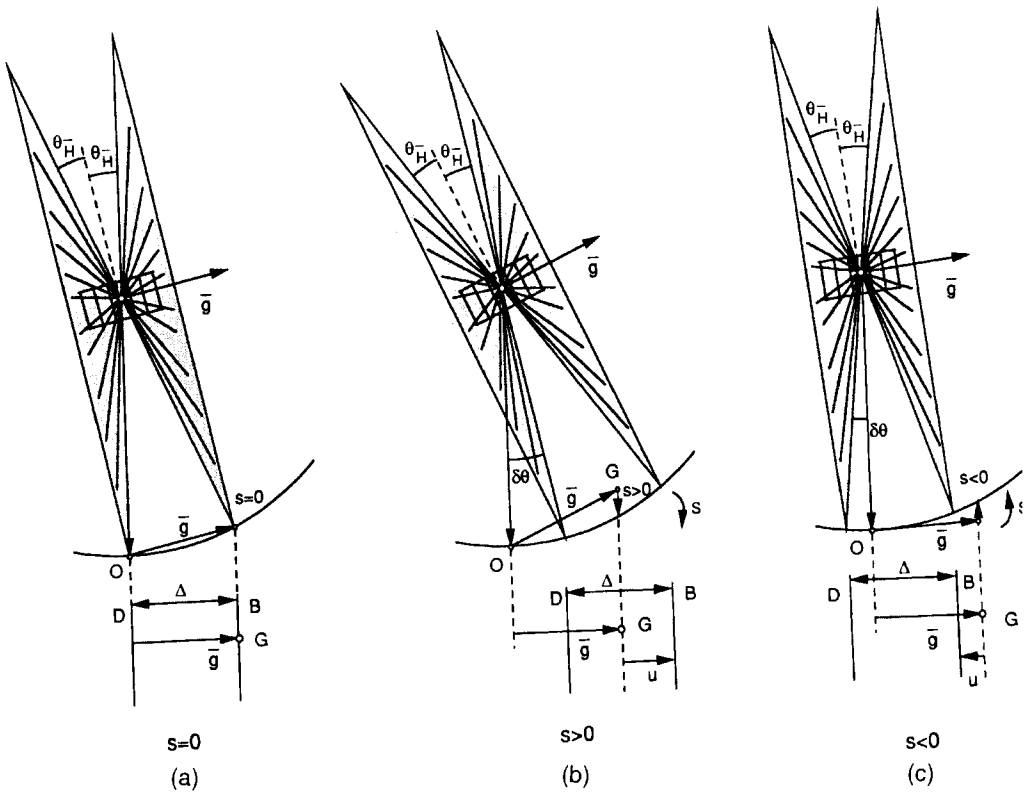


Figure 14. Evolution of the Kikuchi line pattern on tilting. D: dark line; B: bright line; *u*: separation spot-line; Δ: line separation.

sensitive to the orientation, and as a consequence it carries very useful information which can only otherwise be obtained with difficulty, as we shall see.

When the specimen is tilted in such a way that the set of lattice planes ***g*** satisfies the Bragg condition, the situation with respect to the incident beam is no longer symmetrical (Fig. 14). The elastically Bragg scattered beam, which produces the spot G is now one of the generators of the cone. One of the Kikuchi lines thus passes through the Bragg spot. It appears bright (B) on a positive print, that is it corresponds with an excess of electrons above the background. The other line (D) which appears dark due to a deficiency of electrons, passes through the origin. The dark line is produced against a high background caused by the predominantly forward, inelastically scattered electrons. Among these electrons, those which satisfy the Bragg condition are scattered elastically out of this background onto the sheet of the cone which passes through the Bragg spot. Along the parallel line through the origin, which is the locus of the electrons satisfying Bragg's condition, there is as a consequence a deficiency of electrons compared to the background. On the other hand, the same electrons which by their absence cause the dark line through the

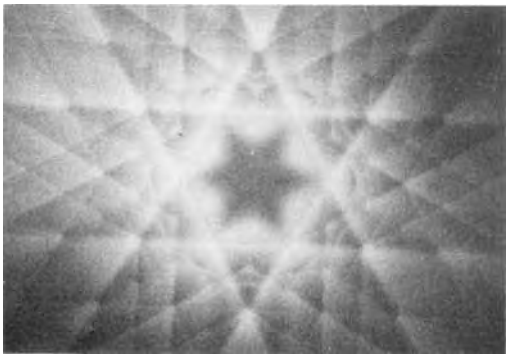


Figure 15. Kikuchi bands in a symmetrically oriented foil of silicon, along the [111] zone.

origin, cause an excess, compared to a lower background, along the part of the cone containing the coherently scattered Bragg beam. This background is somewhat smaller since the scattering angle is larger. Therefore the excess electrons produce a bright line through the Bragg spot. The angular separation of the bright–dark line pair, is the same as in the symmetrical orientation; however, the linear separation measured on the plate may depend slightly on the tilt angle. The symmetrical situation is represented schematically in Fig. 13. In the symmetrical orientation the Kikuchi lines often form the limiting lines of 'Kikuchi bands', the inside of which exhibit a somewhat lower brightness than the outside (Fig. 15) [7, 8]. In this particular orientation the Kikuchi lines can be considered as images of the Brillouin zone boundaries belonging to the different reflections.

Determination of Sign and Magnitude of s

Starting with a foil in the exact Bragg orientation for the reflection G, the bright Kikuchi line passes through G (Fig. 14a), whereas the dark line passes through the origin of the reciprocal lattice. Tilting the specimen over a small angle $\delta\theta$ in the clockwise sense (i.e., towards $s < 0$ about an axis in the foil plane, normal to the g vector), the position of the bright Kikuchi line moves towards the origin over $u = L\,\delta\theta$ (Fig. 14c). The vector $\boldsymbol{g}$ is then rotated over the same angle $\delta\theta$ and hence s becomes negative and equal to $s = g\,\delta\theta$; the relation between u and s is thus:

$$u = \left(\frac{L}{g}\right)s \tag{23a}$$

and

$$\Delta u = \left(\frac{L}{g}\right)\Delta s \tag{23b}$$

This relation allows one to determine the sign and the magnitude of s from the relative position of a diffraction spot and its associated Kikuchi line (Fig. 14). It also allows one to determine the orientation difference between two crystal parts. The sign of s is required for a number of applications such as the determination of the sign of the Burgers vector of dislocation, the vacancy or interstitial character of a dislocation loop, and the orientation difference across a domain boundary, as will be discussed below. The magnitude of s is needed when applying the weak-beam method (see Section 1.1.17.4 of this chapter).

Refraction of Electrons at an Interface

Refraction of the incident electron beam takes place at the vacuum–crystal foil interface because the lengths of the wavevectors are different in the two

media:

$$K_0 \equiv K_{\text{vac}} = \frac{(2meE)^{1/2}}{h} \tag{24}$$

$$K \equiv K_{\text{cryst}} = \frac{[2me(E+V_0)]^{1/2}}{h} \tag{25}$$

but the tangential components have to be conserved at the interface. Figure 16 shows the relation between the two wavevectors; one has

$$\begin{aligned} n &= \sin i / \sin r \\ &= \left(\frac{K_{\text{t,vac}}}{K_0}\right) \Big/ \left(\frac{K_{\text{t,cryst}}}{K_{\text{cryst}}}\right) \\ &= \frac{K_{\text{cryst}}}{K_{\text{vac}}} \\ &= \left(\frac{E+V_0}{E}\right)^{1/2} \end{aligned} \tag{26}$$

The refractive index n is thus:

$$\begin{aligned} n &= [(E+V_0)/E]^{1/2} \\ &\simeq \left(1+\frac{V_0}{E}\right)^{1/2} \end{aligned} \tag{27}$$

Since $V_0 \ll E, n$ is just slightly larger than 1 and the angle of refraction is very small, especially for quasinormal incidence as is the case in most observations. Refraction nevertheless produces an observable effect for grazing incidence. Small polyhedral particles which are completely embedded in the incident beam may produce diffraction spots consisting of a number of components corresponding to the number of crystal wedges crossed by the beam. Refraction also produces an observable effect on the diffraction pattern of single thin fibers such as chrysotile.

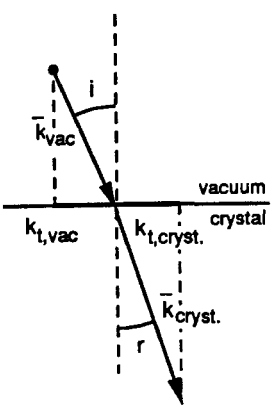


Figure 16. Refraction of electrons at the crystal–vacuum interface.

1.1.1.4 Kinematical Diffraction Theory

Kinematical Diffraction Theory as a Born Approximation

A rigorous diffraction theory is based on Schrödinger's equation, which describes adequately the interaction of the imaging electrons represented by their wave function ψ with the periodic lattice potential of the crystal $V(\boldsymbol{r})$ (a few volts) [9]. The proper equation is thus

$$\left(\frac{h^2}{8\pi^2 m}\right)\Delta\psi + [E+V(\boldsymbol{r})]e\psi = 0 \tag{28}$$

where $-e$ is the charge of the electron, E is the accelerating potential (about 100–400 kV), and the other symbols have their usual meanings. We separate out the constant part V_0 of the lattice potential by introducing $V'(\boldsymbol{r}) = V(\boldsymbol{r}) - V_0$ and $E + V_0 = h^2 k_0^2/2m$. Furthermore, we define the reduced lattice potential $U(\boldsymbol{r})$ as

$$U(\boldsymbol{r}) = \left(\frac{2me}{h^2}\right) V'(\boldsymbol{r}) \tag{29}$$

with $U_0 = 0$ since V_0 was added to E. We expand $U(\boldsymbol{r})$ in a Fourier series:

$$U(\boldsymbol{r}) = \sum_{\boldsymbol{g}} U_{\boldsymbol{g}} \exp(2\pi i \boldsymbol{g} \cdot \boldsymbol{r}) \tag{30}$$

The Schrödinger equation can then be rewritten as

$$\Delta\psi + 4\pi^2 k_0^2 \psi = -4\pi^2 U(\boldsymbol{r})\,\psi \tag{31}$$

A solution of the equation without the right-hand side is the plane wave

$$\psi_0 = C \exp(2\pi i \boldsymbol{k}_0 \cdot \boldsymbol{r}) \tag{32}$$

Equation (32) represents the incident wave with the wavevector $\boldsymbol{k}_0$ corrected for refraction by the constant part V_0 of the lattice potential. The right-hand side of Eq. (31) describes the perturbation caused by the periodic part of the lattice potential which is small compared to V_0. The kinematical approximation, which is equivalent to the first Born approximation, now consists in substituting ψ by ψ_0 in the right-hand side of Eq. (31), and solving the resulting equation:

$$\Delta\psi + 4\pi^2 k_0^2 \psi = -4\pi^2 \sum_{\boldsymbol{g}} U_{\boldsymbol{g}} \exp[2\pi i(\boldsymbol{k}_0 + \boldsymbol{g}) \cdot \boldsymbol{r}] \tag{33}$$

The solution of this equation is of the general form

$$\psi = \psi_0 + \sum_{\boldsymbol{g}} \psi_{\boldsymbol{g}} \tag{34}$$

that is, it consists of the solution of the equation without the right-hand side to which is added a particular solution of the complete equation which is of the same form as the right-hand side. Substituting Eq. (34) in Eq. (33) and remembering that ψ_0 is a solution of the equation without the right-hand side, then $\psi_{\boldsymbol{g}}$ must satisfy the equations

$$\Delta\psi_{\boldsymbol{g}} + 4\pi^2 k_0^2 \psi_{\boldsymbol{g}} = -4\pi^2 U_{\boldsymbol{g}} \exp[2\pi i(\boldsymbol{k}_0 + \boldsymbol{g}) \cdot \boldsymbol{r}] \tag{35}$$

with the boundary condition $\psi_{\boldsymbol{g}} = 0$ at the entrance face $z = 0$. Each equation is seen to involve one beam only; this is consistent with the kinematical approximation which implies that the scattered beams are all independent. Since we are looking for solutions that represent scattered beams, we use an 'ansatz':

$$\psi_{\boldsymbol{g}}(\boldsymbol{r}) = \varphi_{\boldsymbol{g}}(\boldsymbol{r}) \exp[2\pi i(\boldsymbol{k}_0 + \boldsymbol{g}) \cdot \boldsymbol{r}] \tag{36}$$

which has the same form as the right-hand side but an amplitude depending on $\boldsymbol{r}$; it is expected that this dependence on $\boldsymbol{r}$ will be a smooth variation.

Gevers [9] has shown that, substituting this expression in Eq. (35) leads, after some transformations, to

$$\Delta\varphi_{\boldsymbol{g}} + 4\pi i(\boldsymbol{k}_0 + \boldsymbol{g}) \cdot \operatorname{grad} \varphi_{\boldsymbol{g}} + 4\pi^2[k_0^2 + (\boldsymbol{k}_0 + \boldsymbol{g})^2]\varphi_{\boldsymbol{g}} = -4\pi^2 U_{\boldsymbol{g}} \tag{37}$$

With the choice of axis adopted in Fig. 17, and denoting by α the angle between

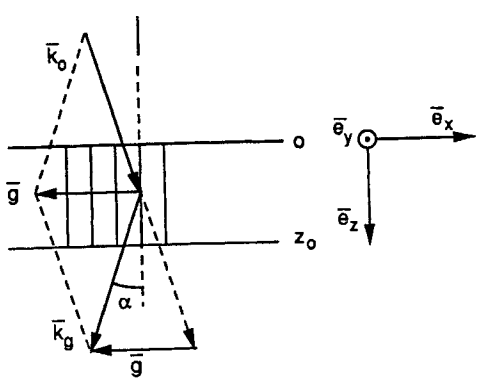


Figure 17. Reference system used in the discussion of the Born approximation.

$\boldsymbol{k}_0 + \boldsymbol{g}$ and $\boldsymbol{e}_z$, we obtain

$$\frac{\partial \varphi_g}{\partial z} + tg\alpha \frac{\partial \varphi_g}{\partial x} - 2\pi i s_g \varphi_g = \frac{\pi i U_g}{|\boldsymbol{k}_0 + \boldsymbol{g}| \cos \alpha} \tag{38}$$

we have hereby neglected the term $\Delta\varphi_g$, which is small compared to the gradient term because φ_g is a smoothly varying function and because the coefficient of the gradient term is large since from geometrical considerations follows that

$$s_g = \frac{k_0^2 - |\boldsymbol{k}_0 + \boldsymbol{g}|^2}{2|\boldsymbol{k}_0 + \boldsymbol{g}| \cos \alpha} \tag{39}$$

The right-hand side of Eq. (38) has the dimensions of inverse length, which still depends on $\boldsymbol{g}$; we shall call it $\pi i / t_g$, where t_g is called the *extinction distance*. Noting that $k_0 \gg \boldsymbol{g}$ we can put $|\boldsymbol{k}_0 + \boldsymbol{g}| \approx k_0$ and obtain

$$t_g = \frac{h^2 k_0 \cos \alpha}{2me|V_g|} \tag{40}$$

and, reintroducing the accelerating voltage E, neglecting $V_0 \ll E$ we obtain

$$t_g = \frac{\lambda E \cos \alpha}{|V_g|} \quad \text{with} \cos \alpha \approx 1 \tag{41}$$

In a plan parallel perfect foil, the scattered amplitude cannot depend on x and hence $\partial \varphi_g / \partial x = 0$. Moreover, even in a deformed foil it is a good approximation to assume that electrons propagate along narrow columns. This 'column' approximation justifies the neglect of the term in $tg\alpha$, which is very small. We then finally obtain as the basic equation for the kinematical theory:

$$\frac{d\varphi_g}{dz} - 2\pi i s \varphi_g = (\pi i / t_g) \exp(i\theta_g) \tag{42}$$

since $V_g = |V_g| \exp(i\theta_g)$.

This equation can readily be integrated between the front ($z = 0$) and exit surface ($z = t_0$) of the foil. Taking into account that $\psi_g = 0$ for $z = 0$, this leads to

$$\varphi_g = (\pi i / t_g)[\exp(2\pi i s t_0)][\exp(i\theta_g)] \times \int_0^{t_0} \exp(-2\pi i s z)\, dz \tag{43}$$

or explicitly, after some simplifications,

$$\varphi_g = i[\exp(\pi i s t_0)][\exp(i\theta_g)][\sin(\pi s t_0 / s t_g)] \tag{44}$$

For the scattered intensity per unit of incident intensity one thus finds

$$I_g = \psi_g \psi_g^* = \varphi_g \varphi_g^* = \frac{\sin^2(\pi s t_0)}{(s t_g)^2} \tag{45}$$

The variation of I_g with thickness is periodic with the period $1/s$, being zero at the entrance face and for $t_0 = n/s$ (whose n is an integer), whereas maxima occur for $t_0 = (n + \frac{1}{2})(1/s)$.

The dependence of I_g on s is represented in Fig. 18. There is a pronounced central maximum (for $s = 0$) given by $I_g(\max) = (\pi t_0 / t_g)^2$. There are zeros for $s = n/t_0$ ($t_0 \equiv z_0$) and maxima approximately halfway between the zeros. These secundary maxima are much smaller than the central peak.

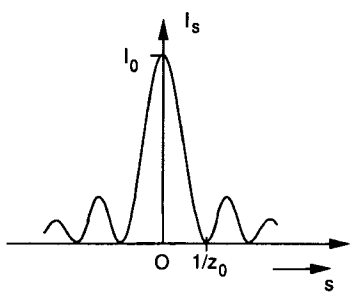


Figure 18. Variation of the diffracted intensity with the deviation parameters (foil thickness = $t_0 \equiv z_0$).

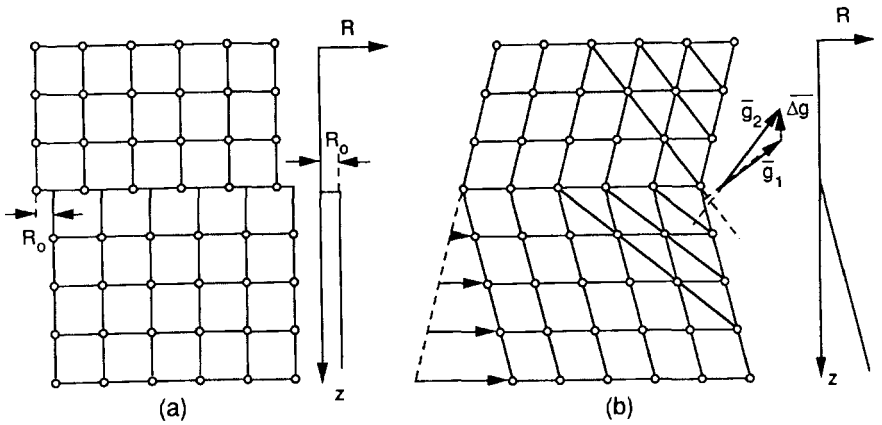


Figure 19. Displacement function of planar interfaces: (a) stacking fault; (b) domain boundary.

Diffraction by Deformed Crystals: Models

Deformations of the crystal can be modeled by stating that the unit cell which in the undeformed crystal was in $\boldsymbol{A}_L$ now occupies, after deformation, the position $\boldsymbol{A}_L + \boldsymbol{R}(\boldsymbol{r})$. The deformation field $\boldsymbol{R}(\boldsymbol{r})$ characterizes the defect. A few simple examples of deformation fields are:

(i) Planar translation interfaces such as stacking faults, out-of-phase boundaries, and discommensuration walls. All these planar defects have a displacement field of the type $\boldsymbol{R} = 0$ for $z \leqslant z_1$ and $\boldsymbol{R} = \boldsymbol{R}_0$ for $z > z_1$, where z_1 is the level at which the planar defect occurs behind the entrance face (Fig. 19a).

(ii) Domain boundaries or twin boundaries with a small twinning vector. We now have $\boldsymbol{R} = 0$ for $z \leqslant z_1$ and $\boldsymbol{R} = \boldsymbol{k}z$ for $z > z_1$ (Fig. 19b).

(iii) A pure screw dislocation has a displacement field described by $\boldsymbol{R} = \boldsymbol{b}[\theta/2\pi]$, where θ is the azimuth angle, measured in the plane perpendicular to $\boldsymbol{b}$. All displacements are clearly parallel to $\boldsymbol{b}$.

(iv) A spherical inclusion has a radial, spherically symmetric displacement field:

$$\boldsymbol{R} = \frac{\varepsilon r_0^3 \boldsymbol{r}}{r^3} \tag{46}$$

for $r \geqslant r_0$; for $r < r_0$, $\boldsymbol{R} = \varepsilon \boldsymbol{r}$ with $\varepsilon = (2/3)\delta$, where δ is the lattice mismatch between inclusion and matrix.

Planar interfaces which are inclined with respect to the foil surface can be considered as consisting of 'steps' that are one column wide. Along a line perpendicular to the intersection line of the fault plane and the foil surface, the columns are assumed to contain a planar fault, parallel to the foil plane, at the level where the inclined fault plane intersects the column. Similarly, an inclined dislocation line is assumed to consist of small segments, each one column long, parallel to the foil plane. It is thus sufficient to consider defects in planes parallel to the foil surfaces.

Scattered Amplitude for a Deformed Foil

Kinematical Formulation

In the case of a deformed foil, the scattered amplitude corresponding to the scattering vector $\boldsymbol{h} = \boldsymbol{g} + \boldsymbol{s}$ becomes, from Eq. (8),

$$A(\boldsymbol{h}) = \sum_L F_g \exp[2\pi i(\boldsymbol{g} + \boldsymbol{s})] \cdot [\boldsymbol{A}_L + R(\boldsymbol{r})] \tag{47}$$

or, replacing the summation by an integration as in Sec. 1.1.3.5 of this Chapter.

$$A(\boldsymbol{h}) = F_g \int_{\text{Column}} \exp\{2\pi i[\boldsymbol{g} \cdot \boldsymbol{R}(\boldsymbol{r}) + sz]\}\, dz \tag{48}$$

Hereby we have used the fact that $\boldsymbol{g} \cdot \boldsymbol{A}_L$ is an integer and that $\boldsymbol{s} \cdot \boldsymbol{R}(\boldsymbol{r})$ is much smaller than the other terms in the exponential. Putting

$$\alpha = 2\pi \boldsymbol{g} \cdot \boldsymbol{R}(\boldsymbol{r}) \tag{49}$$

we can write

$$\boldsymbol{A}(\boldsymbol{h}) = F_g \int_0^{t_0} \exp[i\alpha(z)] \exp[2\pi i s z]\, dz \tag{50}$$

Stacking Fault Contrast

Let the fault plane be parallel with the foil planes at $z = z_1$, behind the entrance face. Since $\boldsymbol{R}$ is constant in this case, the α is also constant and we can split the integral of Eq. (50) into two parts [10]:

$$\frac{A(\boldsymbol{h})}{F_g} = \int_0^{z_1} \exp(2\pi i s z)\, dz + \exp(i\alpha) \int_{z_1}^{t_0} \exp(2\pi i s z)\, dz \tag{51}$$

or, after evaluating the integrals and computing $I(\boldsymbol{h}) = A(\boldsymbol{h})\, A^*(\boldsymbol{h})$

$$\frac{I(\boldsymbol{h})}{F_g^2} = \{1 - \cos(\alpha + \pi s t_0)\cos(\pi s t_0) + \cos(2\pi s u)[\cos(\alpha + \pi s t_0) - \cos \pi s t_0]\}/(\pi s)^2 \tag{52}$$

with $u = \frac{1}{2}(2z_1 - t_0)$, that is u is the distance counted from the central plane of the foil.

The intensity $I(\boldsymbol{h})$ clearly depends periodically on the thickness t_0 of the foil as well as on the level of the fault in the foil (i.e. on u). For an inclined fault in a foil of constant thickness, the intensity $I(\boldsymbol{h})$ is clearly a periodic function of u with period $1/s$; it is symmetrical in u since the cosine is an even function. An electron micrograph will produce a projection of this intensity distribution, that is a fringe pattern with depth period $1/s$, the lateral extent of which is confined to the projected width of the fault (Fig. 20).

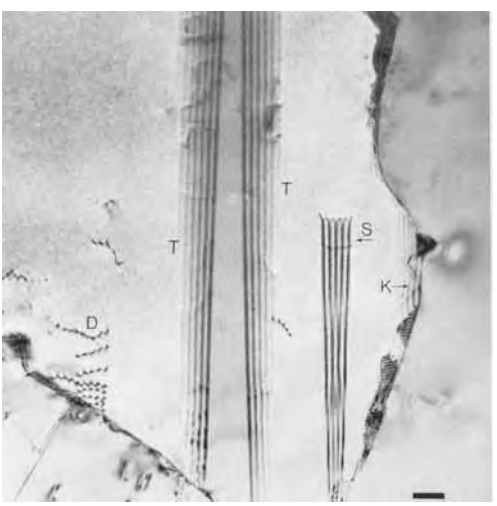


Figure 20. Fringe pattern due to a stacking fault (S), wedge fringes at the twin (T) and dislocations (D) (stainless steel). (Courtesy of J. Van Landuyt.)

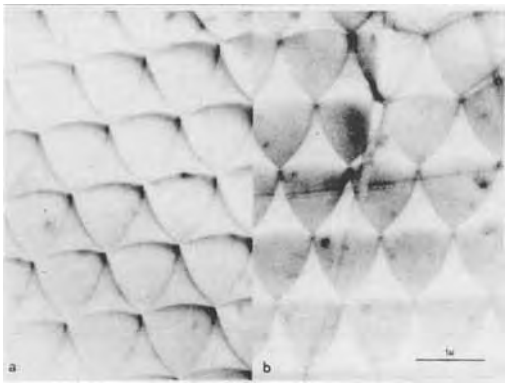


Figure 21. Uniformly shaded area due to the presence of a stacking fault parallel to the foil plane (graphite): (a) partial dislocations in contrast; (b) stacking fault areas show up as bright areas. (Courtesy of P. Delavignette.)

If the fault plane is parallel to the foil surfaces a region of uniform shade is produced in the fault area. This shade can be either brighter or darker than the perfect area of the foil (Fig. 21).

Domain Boundary Contrast

Let the domain boundary be parallel to the foil surfaces and situated at z_1 (Fig. 22). We then have $\alpha = 0$ for $z \leqslant z_1$ and $\alpha = 2\pi \boldsymbol{g} \cdot \boldsymbol{k} z$ for $z_1 < z \leqslant t_0$. The integral [Eq. (50)] can again be split into two parts. Recalling that $\boldsymbol{g} \cdot \boldsymbol{k} = \Delta s$, we have

$$\frac{A(\boldsymbol{h})}{F_h} = \int_0^{z_1} \exp(2\pi i s z)\, dz + \int_{z_1}^{z_0} \exp[2\pi i (s + \Delta s) z]\, dz \qquad (53)$$

Since the two crystal parts on both sides of the surface are perfect, but slightly misoriented, s and Δs are constant. These integrals can easily be evaluated explicitly.

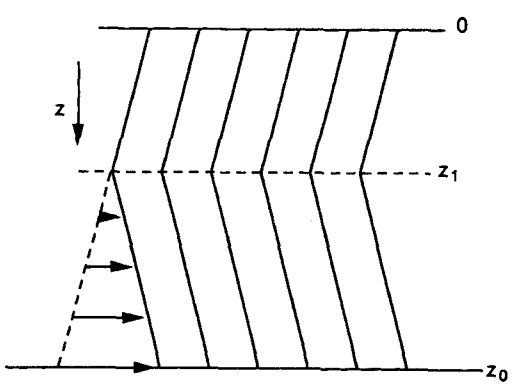


Figure 22. Model for a domain boundary.

Dislocation Contrast

We adopt the geometry shown in Fig. 23, that is the dislocation line is parallel to the foil plane (x, y), with the dislocation lying along the y axis [11]. We then have $\theta = \text{arctg}(z/x)$ and, since $\boldsymbol{R} = \boldsymbol{b}(\theta/2\pi)$, from Eq. (49)

$$\alpha = (\boldsymbol{g} \cdot \boldsymbol{b}) \,\text{arctg}\left(\frac{z}{x}\right) \qquad (54)$$

where $\boldsymbol{g} \cdot \boldsymbol{b} = n$ is an integer for a perfect dislocation but a fraction for a partial dislocation. The scattered amplitude is given by

$$\frac{A(\boldsymbol{h})}{F_g} = \int_{-z_1}^{+z_2} \exp(2\pi i s z) \cdot \exp\left[in \,\text{arctg}\left(\frac{z}{x}\right)\right] dz \qquad (55)$$

The integral can be evaluated for different values of n but the results are not

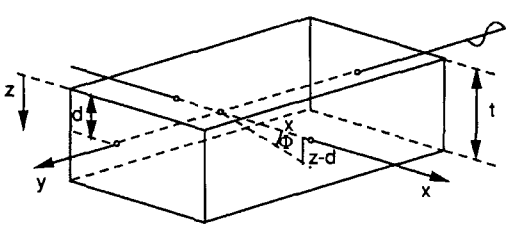


Figure 23. Foil containing a screw dislocation; the notation used.

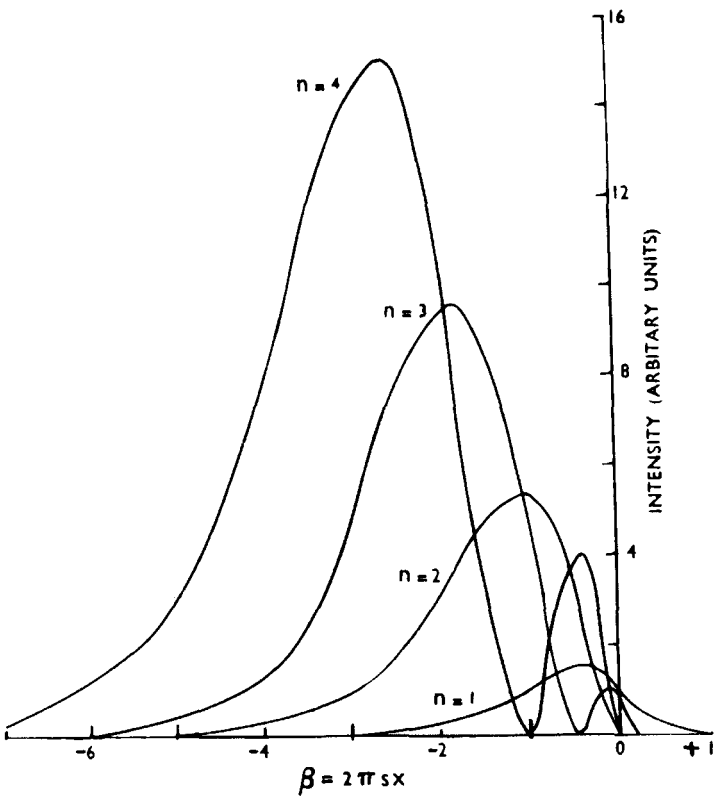


Figure 24. Image profiles for screw dislocations according to the kinematical theory [13, 14].

elementary functions. The numerical results are represented graphically in Fig. 24 for different values of n, as a function of $\beta = 2\pi sx$. It turns out that the peak height, peak shift and peak width increase with increasing n [11]. It is clear that no image will be produced if $n = 0$ because then Eq. (55) reduces to that for a perfect foil. The corresponding profiles for edge dislocations are represented in Fig. 25 [12].

Lattice Potential of Deformed Crystals

Deformation can also be introduced in the diffraction equations using the 'deformable ion approximation' via the lattice potential [13, 14]. It is assumed that a displaced volume element of crystal carries along with it the lattice potential it had before the deformation

$$V_{\text{def}}(\boldsymbol{r}) = V_{\text{undef}}(\boldsymbol{r} - \boldsymbol{R}) \tag{56}$$

As a result the Fourier coefficients of the lattice potential in the deformed crystal acquire phase factors since

$$\begin{aligned} V_{\text{def}}(\boldsymbol{r}) &= V_0 + \sum_{\boldsymbol{g}} V_{\boldsymbol{g}} \exp[2\pi i \boldsymbol{g} \cdot (\boldsymbol{r} - \boldsymbol{R})] \\ &= V_0 + \sum_{\boldsymbol{g}} [V_{\boldsymbol{g}} \exp(-2\pi i \boldsymbol{g} \cdot \boldsymbol{R})] \\ &\quad \times \exp(2\pi i \boldsymbol{g} \cdot \boldsymbol{r}) \end{aligned} \tag{57}$$

Strictly speaking, this series is no longer a Fourier series, which is consistent with the fact that the function $V_{\text{def}}(\boldsymbol{r})$ is no longer a

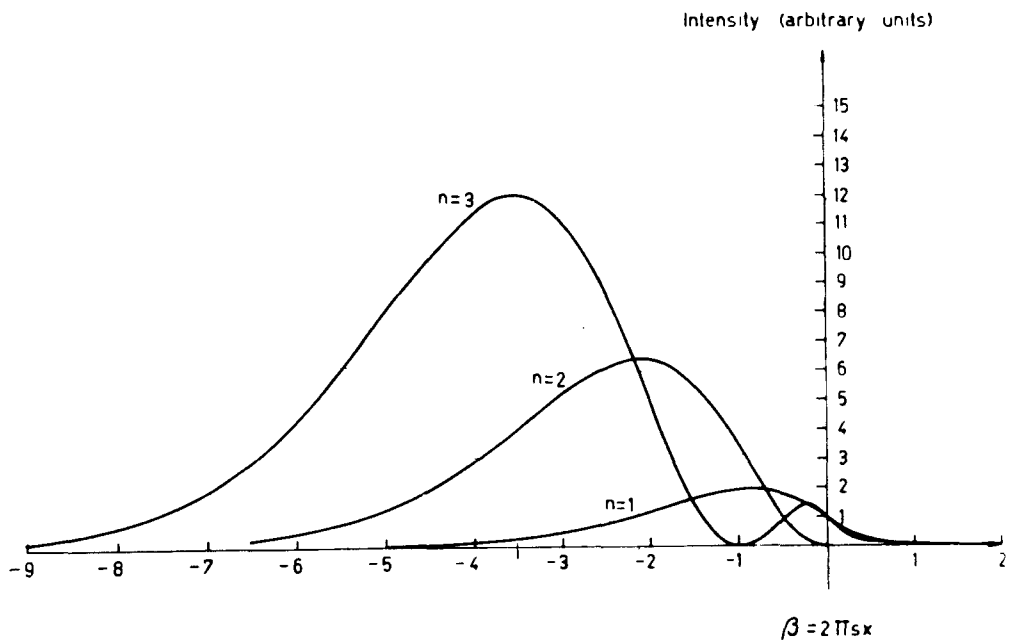


Figure 25. Image profiles for edge dislocations according to the kinematical theory [12].

periodic function. Accepting nevertheless this approximation, the presence of a deformation field can be accounted for by replacing V_g by

$$V_g \rightarrow V_g \exp[-2\pi i \boldsymbol{g} \cdot \boldsymbol{R}(\boldsymbol{r})] \tag{58}$$

In view of the relations between V_g and $1/t_g$ this also amounts to replacing $1/t_g$ by $(1/t_g)\exp(-i\alpha_g)$, with $\alpha_g = 2\pi \boldsymbol{g} \cdot \boldsymbol{R}(\boldsymbol{r})$.

For a deformed foil, Eq. (42) thus becomes

$$\frac{\mathrm{d}\phi_g}{\mathrm{d}z} - 2\pi i s_g \phi_g = \left(\frac{\pi i}{t_g}\right)\exp(i\theta_g)\exp(-i\alpha_g) \tag{59}$$

Introducing a new function $\varphi_g = \phi_g \exp(-i\alpha_g)$ which does not change the scattered intensity $I_g = \varphi_g \varphi_g^* = \phi_g \phi_g^*$ this equation becomes

$$\frac{\mathrm{d}\varphi_g}{\mathrm{d}z} - 2\pi i\left(s_g + \boldsymbol{g} \cdot \frac{\mathrm{d}\boldsymbol{R}}{\mathrm{d}z}\right)\varphi_g = \left(\frac{\pi i}{t_g}\right)\exp(i\theta_g) \tag{60}$$

This equation is of the same form as Eq. (42), except that s_g has been replaced by a local value

$$s_{\mathrm{eff}} = s_g + \boldsymbol{g} \cdot \frac{\mathrm{d}\boldsymbol{R}}{\mathrm{d}z} \tag{61}$$

which is, in general, a function of x, y, and z.

Depending on the problem to be treated, one has to use Eq. (60) or Eq. (59); if α_g is a constant, Eq. (60) cannot be used since it reduces to that of a perfect crystal.

1.1.1.5 Two-Beam Dynamical Theory [13–15]

Basic Equations for Perfect Crystals

Rather than directly deriving the fundamental equations of the dynamical theory from Schrödinger's equation, we shall use a physically transparent method, related to the one originally used by Darwin [16]. We introduce the usual approximations (high energy approximation, column approximation, etc.) already in the model, rather than in the final equations.

The electrons traveling down a column of crystal can be represented by the wavefunction:

$$\psi(\boldsymbol{r}) = \phi_0(z)\exp(2\pi i\boldsymbol{k}_0\cdot\boldsymbol{r}) + \phi_g(z)\exp(2\pi i\boldsymbol{k}\cdot\boldsymbol{r}) \qquad (62)$$

where the $\boldsymbol{k}$ and $\boldsymbol{k}_0$ represent the wavevectors of the scattered and of the incident beam, respectively; the amplitudes ϕ_g and ϕ_0 of these beams depend on z. The column approximation justifies to neglect the x dependence of these amplitudes. We shall write down the amplitude after interaction with a slice dz of the column, at level z behind the entrance face (Fig. 26).

We can write

$$\phi_0(z+\mathrm{d}z) = \phi_0(z)\phi_0(\mathrm{d}z) + \phi_g(z)\phi_{-g}(\mathrm{d}z) \qquad (63)$$

where ϕ_{-g} means that scattering with scattering vector $-\boldsymbol{g}$ has to be taken into account. This relation states that the transmitted beam results from the interference between the twice transmitted beam of which the amplitude is $\phi_0(z)\,\phi_0(\mathrm{d}z)$ and the twice-scattered beam with amplitude $\phi_g(z)\,\phi_{-g}(\mathrm{d}z)$. Since the slices d$z$ are arbitrarily thin, it is justified to use the kinematical approximation for $\phi_0(\mathrm{d}z) = 1$ and $\phi_g(\mathrm{d}z)$. Note that changing $\boldsymbol{g} \rightarrow -\boldsymbol{g}$ implies changing $s \rightarrow -s$, since the origin of reciprocal space is now in the node point G. From Eq. (43) we can conclude that

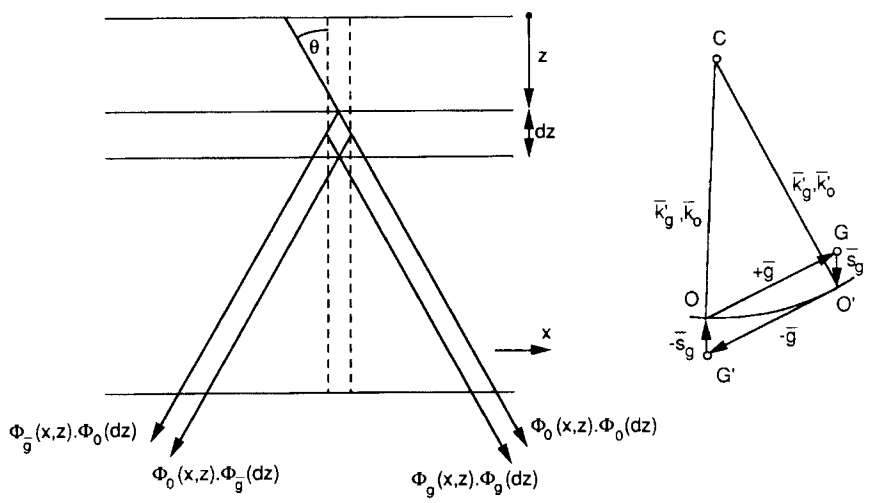


Figure 26. Illustrating the derivation of the fundamental equations of the dynamical theory.

(ignoring the phase factor of the Fourier coefficient)

$$\phi_g(\mathrm{d}z) = \mathrm{d}\phi_g = \left(\frac{\pi \mathrm{i}}{t_g}\right)\exp(-2\pi \mathrm{i}sz)\,\mathrm{d}z \tag{64}$$

Furthermore, we have $\phi_0(\mathrm{d}z) = 1$ and hence

$$\phi_0(z + \mathrm{d}z) = \phi_0(z) + \phi_g(z)\left(\frac{\pi \mathrm{i}}{t_{-g}}\right) \times \exp(2\pi \mathrm{i}sz)\,\mathrm{d}z \tag{65}$$

or

$$\phi_0(z + \mathrm{d}z) - \phi_0(z) = \phi_g(z)\left(\frac{\pi \mathrm{i}}{t_{-g}}\right)\exp(2\pi \mathrm{i}sz)\,\mathrm{d}z \tag{66}$$

and in the limit for $\mathrm{d}z \to 0$:

$$(\mathrm{d}\phi_0/\mathrm{d}z) = \left(\frac{\pi \mathrm{i}}{t_{-g}}\right)[\exp(2\pi \mathrm{i}sz)]\phi_g(z) \tag{67}$$

Similarly, for the scattered beam we can write

$$\phi_g(z + \mathrm{d}z) = \phi_0(z)\phi_g(\mathrm{d}z) + \phi_g(z)\,\phi_0(\mathrm{d}z) \tag{68}$$

and taking into account the expressions for $\phi_g(\mathrm{d}z)$ [Eq. (64)] and $\phi_0(\mathrm{d}z) = 1$, and making the same transformations as for ϕ_0, we obtain:

$$\mathrm{d}\phi_g/\mathrm{d}z = \left(\frac{\pi \mathrm{i}}{t_g}\right)[\exp(-2\pi \mathrm{i}sz)]\phi_0(z) \tag{69}$$

Equations (67) and (69) form the Darwin–Howie–Whelan set of coupled differential equations describing the interplay between incident and scattered beams in a perfect crystal. This set can be given a number of alternative forms by introducing amplitudes differing only by phase factors from ϕ_0 and ϕ_g. This does not change the resulting intensity distributions.

One can, for instance, substitute T' and S' for ϕ_0 and ϕ_g

$$\phi_0 = T' \exp(\pi \mathrm{i}sz) \tag{70a}$$

$$\phi_g = S' \exp(-\pi \mathrm{i}sz) \tag{70b}$$

The system of equations then acquires a symmetrical form

$$\mathrm{d}T'/\mathrm{d}z + \pi \mathrm{i}sT' = \left(\frac{\pi \mathrm{i}}{t_{-g}}\right)S' \tag{71a}$$

$$\mathrm{d}S'/\mathrm{d}z - \pi \mathrm{i}sS' = \left(\frac{\pi \mathrm{i}}{t_g}\right)T' \tag{71b}$$

or one can make the substitution

$$\phi_0 = T \tag{72a}$$

$$\phi_g = S \exp(-2\pi \mathrm{i}sz) \tag{72b}$$

and obtain the following asymmetrical set of equations:

$$\mathrm{d}T/\mathrm{d}z = \left(\frac{\pi \mathrm{i}}{t_{-g}}\right)S \tag{73a}$$

$$\mathrm{d}S/\mathrm{d}z = 2\pi \mathrm{i}sS + \left(\frac{\pi \mathrm{i}}{t_g}\right)T \tag{73b}$$

Depending on the problem to be solved, either the symmetrical or the asymmetrical set will have to be used.

Dynamical Equations for Deformed Crystals

We have seen previously (Sec. 1.1.4.4 of this Chapter) that the deformation of the lattice can be modeled by using a local effective deviation parameter, s_{eff}, which is a function of $\boldsymbol{r}$, rather than a constant s.

We found that s_{eff} is given by

$$s_{\text{eff}} = s + \frac{\mathrm{d}}{\mathrm{d}z}(\boldsymbol{g} \cdot \boldsymbol{R}) \tag{74a}$$

$$s_{\text{eff}} = s + \frac{\mathrm{d}\alpha}{\mathrm{d}z} \tag{74b}$$

with $\alpha = \alpha_g/2\pi$. One can alternatively replace $1/t_g$ by $1/t_g[\exp(-\mathrm{i}\alpha_g)]$. This can be shown directly by intuitive considerations [17].

These substitutions allow us to adapt the different sets of equations to the case of deformed crystals. The presence of a defect is now described by the variation of s along the integration columns, the diffraction vector $\boldsymbol{g}$ being the same all along the column.

Solution of the Two-beam Dynamical Equations for a Perfect Crystal

The two equations of the system [Eqs. (73a) and (73b)] can be uncoupled by eliminating S and T, respectively. For S, one obtains for instance

$$\frac{\mathrm{d}^2 S}{\mathrm{d}z^2} - 2\pi \mathrm{i} s_g \frac{\mathrm{d}S}{\mathrm{d}z} + \frac{\pi^2}{t_g^2} \cdot S = 0 \tag{75}$$

and similarly for an equation for T.

Taking into account the initial values $T = 1$, $S = 0$ for $z = 0$, and requiring that the solutions satisfy both equations of the set, one finds

$$T = \left[\cos(\pi\sigma_g z) - \mathrm{i}\left(\frac{s_g}{\sigma_g}\right)\sin(\pi\sigma_g z)\right] \times \exp(\pi \mathrm{i} s_g z) \tag{76a}$$

$$S = (\mathrm{i}/\sigma_g t_g)[\sin(\pi\sigma_g z)\exp(\pi \mathrm{i} s_g z)] \tag{76b}$$

with

$$\sigma_g^2 = \frac{(1 + s^2 t_g^2)}{t_g^2} \tag{77}$$

The scattered intensity I_{S} is then

$$I_{\text{S}} = SS^* = \sin^2\left[\frac{\pi\sigma_g z}{(\sigma t_g)^2}\right] \tag{78}$$

and

$$I_{\text{T}} = 1 - I_{\text{S}} \tag{79}$$

since we have, so far, neglected absorption.

The depth periods of both I_{S} and I_{T} are now $1/\sigma_g$ as compared to $1/s_g$, according to the kinematical theory. For $s_g = 0$ the depth period remains finite and equal to t_g. Note that for large values of s_g, σ_g reduces to s_g.

Equations (78) and (79) as well as Eqs. (67) and (69) describe the periodic transfer of electrons from the transmitted beam into the scattered beam, and vice versa (Pendellösung effect). The periodic variations of I_{S} (or I_{T}) with the crystal thickness give rise to the formation of thickness extinction contours in wedge-shaped crystals (Fig. 27). The contours are the geometrical loci of equal thickness. According to Eqs. (78) and (79) such fringes should be periodic; in actual fact the fringes are found to be damped with increasing thickness, showing that absorption takes place. We take absorption into account below.

The expressions I_{S} and I_{T} also describe the dependence on s_g. The geometrical loci with $s_g = \text{Constant}$ are the equi-inclination or bent contours. According to the expressions for I_{S} and I_{T}, this dependence should be symmetrical in s, i.e. $I_{\text{S}}(-s) = I_{\text{S}}(+s)$ and $I_{\text{T}}(-s) = I_{\text{T}}(+s)$. In fact, one finds that $I_{\text{T}}(-s) < I_{\text{T}}(+s)$, whereas $I_{\text{S}}(-s) = I_{\text{S}}(+s)$. This effect, called

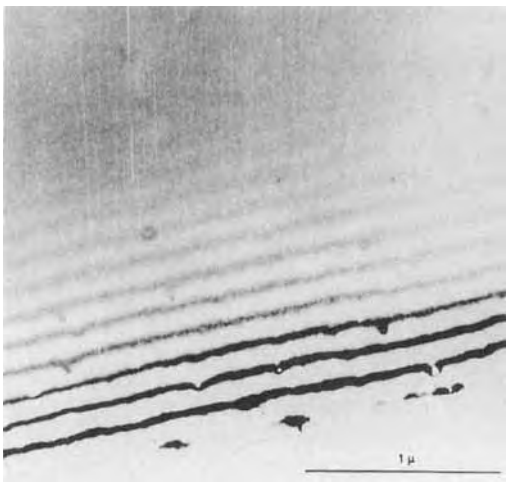


Figure 27. Thickness contours in a silicon wedge.

the *Borrmann effect*, is also a consequence of anomalous absorption (see below).

Two-beam Lattice Fringes

In the two-beam case the total wavefunction at the exit face of a crystal foil can be written as [18–20]

$$\Psi = \exp(2\pi i \boldsymbol{K} \cdot \boldsymbol{r})[T + S\exp(2\pi i \boldsymbol{g} \cdot \boldsymbol{r})] \tag{80}$$

where $\boldsymbol{K}$ is the wavevector of the incident electron beam corrected for refraction by the constant part, V_0, of the lattice potential. The total wavefunction Ψ can be imaged in the microscope by collecting the two interfering beams in the objective aperture and focusing on the exit face of the foil. One then observes the intensity distribution

$$\begin{aligned} I &= \Psi\Psi^* \\ &= TT^* + SS^* + T^*S\exp(2\pi i \boldsymbol{g} \cdot \boldsymbol{r}) \\ &\quad + TS^*\exp(-2\pi i \boldsymbol{g} \cdot \boldsymbol{r}) \qquad (81) \\ &= I_T + I_S + 2\sqrt{I_T I_S}\sin(2\pi i \boldsymbol{g} \cdot \boldsymbol{r} + \varphi) \qquad (82) \end{aligned}$$

where

$$\operatorname{tg}\varphi = \frac{s_g}{\sigma_g}\operatorname{tg}(\pi\sigma_g z_0) \tag{83}$$

z_0 being the foil thickness. The image consists of sinusoidal fringes with a period $1/|\boldsymbol{g}|$, which is equal to the interplanar distance of the lattice planes normal to $\boldsymbol{g}$. The fringes are parallel to the traces of these lattice planes. For $s_g = 0$, Eq. (82) simply reduces to

$$I = 1 + \sin\left(\frac{2\pi z_0}{t_g}\right)\sin(2\pi g \cdot x) \tag{84}$$

where the x axis was chosen along $\boldsymbol{g}$.

The contrast of the fringes clearly depends on z_0, whereas the localization of the fringes depends on s_g, since this determines φ. For $s = 0$, $\varphi = 0$, the bright fringes coincide with the lattice planes $gx =$ Integer $+ \frac{1}{2}$. The fringes are therefore called *lattice fringes*; their formation can be understood on a purely geometrical basis [21].

The phase φ in Eq. (82) will, in general, also depend on the exact imaging conditions. Since the beams T and S may enclose different angles with the optical axis of the microscope, additional phase shifts will, in general, occur. As long as we consider diffraction contrast images, the image transfer function of the microscope is not important since the intensity distribution in a single beam does not depend on its shape. However, when considering the interference of more than one beam, the function plays an important role and cannot be ignored.

1.1.1.6 Absorption

Absorption can be described phenomenologically by assuming the refractive index

$n = c/v$ (where c is the velocity in a vacuum, and v is the velocity in a medium) to be complex. Representing a plane wave as

$$\psi = \psi_0 \exp[\mathrm{i}(kz - \omega t)] \tag{85}$$

with $k = \omega/v = \omega n/c$ and making n complex $n_c = n + \mathrm{i}\mu$ leads to a complex wavevector

$$k_c = \frac{\omega}{c}(n + \mathrm{i}\mu) \tag{86}$$

and to a wave

$$\psi = \psi_0 \exp[\mathrm{i}(k_c z - \omega t)]$$
$$= \psi_0 \exp(-\omega\mu z/c) \exp[\mathrm{i}(kz - \omega t)] \tag{87}$$

which is clearly damped, the amplitude absorption coefficient being $\mu\omega/c$.

Replacing V_0 by $V_0 + \mathrm{i}W_0$ in the lattice potential has the desired effect, since it makes the wavevector K complex:

$$K^{*2} = \frac{2me[(E + V_0) + \mathrm{i}W_0]}{h^2} \tag{88}$$

$$= K^2 + \mathrm{i}[2meW_0/h^2K]K \tag{89}$$

We now define

$$\frac{1}{\tau_0} = 2meW_0/h^2K \tag{90}$$

by analogy with

$$\frac{1}{t_0} = 2meV_0/h^2K \tag{91}$$

The complex wavevector K^* then becomes

$$K^{*2} = K^2 + \mathrm{i}\left(\frac{K}{\tau_0}\right) \tag{92}$$

and, since $1/K\tau_0 \ll 1$, then

$$K^* = K\left(1 + \frac{\mathrm{i}}{K\tau_0}\right)^{1/2}$$
$$\simeq K\left(1 + \frac{\mathrm{i}}{2K\tau_0}\right)$$

or, finally,

$$K^* = K + \frac{\mathrm{i}}{2\tau_0} \tag{93}$$

The boundary conditions now require that at the entrance face $z = 0$ the tangential component must be conserved, i.e. $K_t^* = K_t$. The imaginary part being zero for $z < 0$, this means that for $z > 0$ the imaginary part must be oriented along $\boldsymbol{e}_n$, i.e.

$$\boldsymbol{K}^* = \boldsymbol{K} + \left(\frac{\mathrm{i}}{2\tau_0}\right)\boldsymbol{e}_n \tag{94}$$

The expression $\exp(2\pi\mathrm{i}\boldsymbol{K}\cdot\boldsymbol{r})$ then becomes

$$\exp(2\pi\mathrm{i}\boldsymbol{K}\cdot\boldsymbol{r}) \exp\left(\frac{-\pi z}{\tau_0}\right) \tag{95}$$

since $\boldsymbol{e}_n \cdot \boldsymbol{r} = z$. The absorption coefficient for the amplitude is thus $\mu = \pi/\tau_0$.

It has been shown by Yosioka [22] that anomalous absorption can be taken into account by assuming the lattice potential to become complex: $V(\boldsymbol{r}) + \mathrm{i}W(\boldsymbol{r})$. We have shown that this applies to the constant term $V_0 + \mathrm{i}W_0$, and leads to normal absorption in that case. This procedure can be generalized by replacing also the other Fourier coefficients by complex quantities: $V_g \rightarrow V_g + \mathrm{i}W_g$. In view of the relation between $1/t_g$ and V_g [Eq. (41)], this is equivalent to replacing

$$\frac{1}{t_g} \rightarrow \frac{1}{t_g} + \frac{\mathrm{i}}{\tau_g} \tag{96}$$

where we have introduced the *absorption lengths*, τ_g, which are related to W_g in the same way as t_g are related to V_g:

$$\frac{1}{\tau_g} = \frac{2meW_g}{h^2k} \tag{97}$$

$\tau_g \gg t_g$, since $W_g \ll V_g$. If $1/t_g$ becomes complex it is clear that σ_g [defined in

Eq. (77)] must also become complex

$$\sigma_g = \sigma_r + i\sigma_i \tag{98}$$

Omitting, for simplicity, the index $\boldsymbol{g}$ at σ_r and σ_i, one finds a good approximation, noting that $\sigma_i^2 \ll \sigma_r^2$ and $1/\tau_g^2 \ll 1/t_g^2$:

$$\sigma_r = (1/t_g)[1 + s_g^2 t_g^2]^{1/2} \tag{99a}$$

$$\sigma_i = (\sigma_r t_g \tau_g)^{-1} \tag{99b}$$

We note that $\sigma_i \cdot \boldsymbol{r} = \sigma_i z$.

1.1.1.7 Dynamical Equations Including Absorption

Anomalous absorption effects can be taken into account in the basic equations [Eqs. (67) and (69)] or in the equations derived from these, by the above-mentioned substitutions [Eqs. (96) or (98)]. Equations (67) and (69) then become, for instance,

$$\frac{d\phi_0}{dz} = \pi i\left[\frac{1}{t_{-g}} + i\left(\frac{1}{\tau_{-g}}\right)\right]\exp(2\pi i s_g z)\,\phi_g(z) \tag{100}$$

$$\frac{d\phi_g}{dz} = \pi i\left[\frac{1}{t_g} + i\left(\frac{1}{\tau_g}\right)\right]\exp(-2\pi i s_g z)\,\phi_0(z) \tag{101}$$

The substitution can also simply be performed directly in the final solutions [Eq. (76)]. Although the method is purely phenomenological, the result allows a physically meaningful interpretation. Moreover, in order to obtain tractable analytical solutions one adopts the following approximation: σ is replaced by $\sigma_r + i\sigma_i$ in the periodic terms of the Eqs. (76a) and (76b), but in the coefficients we replace σ by σ_r. This approximation preserves the essential features of anomalous absorption. It allows to show that the wavefunction, ψ_T, of the transmitted beam results from the interference of two waves with slightly different wavevectors ($\boldsymbol{e}$ = unit normal to the foil)

$$\boldsymbol{K} + \tfrac{1}{2}(s + \sigma_r)\boldsymbol{e} \tag{102a}$$

and

$$\boldsymbol{K} + \tfrac{1}{2}(s - \sigma_r)\boldsymbol{e} \tag{102b}$$

For the wavefunction, ψ_S, of the scattered beam, the two interfering waves have wavevectors

$$\boldsymbol{K} + \boldsymbol{g} + \tfrac{1}{2}(s + \sigma_r)\boldsymbol{e} \tag{103a}$$

and

$$\boldsymbol{K} + \boldsymbol{g} + \tfrac{1}{2}(s - \sigma_r)\boldsymbol{e} \tag{103b}$$

The beating of these two waves causes the periodic depth variation of ψ_T and ψ_S with a period $1/\sigma_r$, that is the Pendellösung effect.

The four waves present in the total wavefunction, $\psi = \psi_T + \psi_S$, belong to two Bloch wavefields, ψ_1 and ψ_2. For the waves of the first wavefield, ψ_1, the sign of σ_r is positive; it contains the two waves with wavevectors $\boldsymbol{K} + \frac{1}{2}(s + \sigma_r)\boldsymbol{e}$ and $\boldsymbol{K} + \boldsymbol{g} + \frac{1}{2}(s + \sigma_r)\boldsymbol{e}$. The second wavefield, ψ_2, corresponds to the negative sign of σ_r; it contains the two waves with wavevectors $\boldsymbol{K} + \frac{1}{2}(s - \sigma_r)\boldsymbol{e}$ and $\boldsymbol{K} + \boldsymbol{g} + \frac{1}{2}(s - \sigma_r)\boldsymbol{e}$. In the simple case $s = 0$, the wavefield ψ_1 has a maximum amplitude in planes coinciding with the atomic planes, whereas ψ_2 has maximum amplitude along planes exactly in between the atomic planes [13, 14].

The first type of wavefield which is strongly excited for $s < 0$ is strongly attenuated, whereas the second wavefield, which is more strongly excited for $s > 0$,

is less attenuated by anomalous absorption. These results are physically meaningful. Electrons propagating close to atomic cores, that is electrons for which the wavefunction is peaked at the atomic positions, will have a larger probability to excite X-rays and hence to be 'absorbed' than will electrons passing between atomic planes.

The damping of the thickness fringes in a wedge-shaped crystal can be understood on the basis of these results. The depth variations of I_S and I_T are in both cases caused by the beating of two waves, one of each wavefield. Since one of these waves is attenuated much more than the other, the beating envelope (which has a maximum amplitude when the two beating waves have equal amplitude) decreases with depth in the crystal, even though one of the waves has still an appreciable amplitude.

The Borrmann effect [23] can also be understood on the basis of the same model. For $s > 0$ the amplitude of the rapidly attenuated wave in ψ_T, that is $\frac{1}{2}[1 - (s/\sigma_2)]$, is smaller than that of the passing wave, that is $(\frac{1}{2}[1 + (s/\sigma_2)])$, which is enhanced by anomalous absorption. As a result ψ_T will have a larger amplitude for $s > 0$ than for $s < 0$ for the same absolute value of s. A similar asymmetry is absent for ψ_S, where the amplitudes of the two constituent waves are both $1/(2\sigma t_g)$, which only depends on s^2.

1.1.1.8 Rocking Curves for Perfect Crystals Taking into Account Anomalous Absorption

Explicit expressions for I_T and I_S are obtained by computing $I_T = \psi_T \cdot \psi_T^*$ and $I_S = \psi_S \cdot \psi_S^*$. Since $\sigma_i \ll \sigma_r$ and $\tau_g \gg t_g$ one can approximate the expressions by neglecting higher order terms in t_g/τ_g. After lengthy but straightforward calculations, one obtains

$$I_T = [\cosh u + (s/\sigma_r) \sinh u]^2 - \left[\frac{1}{(\sigma_r t_g)^2}\right] \sin^2 v \tag{104}$$

with $u = \pi\sigma_i z$, $v = \pi\sigma_r z$, $\sigma_i = (\sigma_r t_g \tau_g)^{-1}$, and $\sigma_r = [1 + (st_g)^2]^{1/2}/t_g$. Similarly,

$$I_S = \frac{\sinh^2 u + \sin^2 v}{(\sigma_r t_g)^2} \tag{105}$$

These expressions are represented in Fig. 28 for a rather thin foil. Note the asymmetry of I_T with respect to s (Borrmann effect), even for a rather thin foil.

In the limit $\sigma_i \to 0$, that is as $u \to 0$, one obtains the corresponding expressions for the non-absorption case:

$$I_S = \frac{\sin^2 \pi\sigma_r z}{(\sigma_r t_g)^2} \tag{106a}$$

and

$$I_T = 1 - I_S \tag{106b}$$

For $s = 0$ one obtains

$$I_S = \sinh^2 u + \sin^2 v \tag{107a}$$

and

$$I_T = \cosh^2 u - \sin^2 v \tag{107b}$$

Note that now

$$I_T + I_S = \sinh^2 u + \cosh^2 u > 1!$$

This apparently absurd result is due to the fact that *normal* absorption, which attenuates both beams to the same extent, has been neglected. It can readily be introduced by multiplying by $\exp(-\pi z_0/\tau_0)$ [from Eq. (95)].

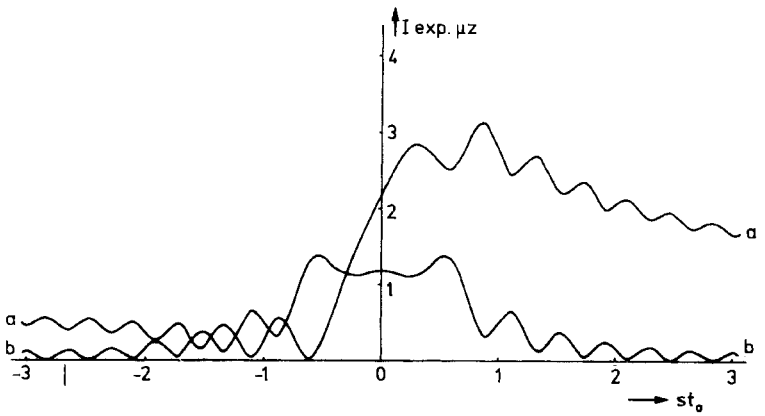


Figure 28. Rocking curves for a perfect crystal taken into account in anomalous absorption. The transmitted intensity I_T (a) is asymmetrical in $s \cdot t_g$, whereas the scattered intensity I_S (b) is symmetrical in $s \cdot t_g$ [12].

The expressions for I_S and I_T can be rewritten ($s = 0$) as

$$I_{T,S} = \tfrac{1}{2}[\cosh 2u \pm \cos 2v] \tag{108}$$

where the plus sign refers to the transmitted beam and the minus sign refers to the scattered beam.

1.1.1.9 Dynamical Diffraction by Deformed or Faulted Crystals [13, 14, 24]

Basic Equations

Quantitative studies of defects in crystals are usually performed under optimized two-beam conditions, since this allows the most straightforward and detailed interpretation. The system of two-beam equations which lends itself most conveniently to the description of diffraction by non-perfect crystals is the Darwin–Howie–Whelan system of coupled differential equations.

Different formulations of this system are available; their solutions for the amplitudes of scattered (S) and transmitted (T) beams differ by phase factors only, and hence lead to the same intensity distribution, that is to the same image. We shall use two different forms of this system, depending on the problem to be treated.

The asymmetric form [from Eq. (73)]:

$$\frac{dT}{dz} = \left(\frac{\pi i}{t_{-g}}\right) S \tag{109a}$$

$$\frac{dS}{dz} = 2\pi i s_g S + \left(\frac{\pi i}{t_g}\right) T \tag{109b}$$

and the symmetric form [from Eq. (71)]

$$\frac{dT'}{dz} + \pi i s_g T' = \left(\frac{\pi i}{t_{-g}}\right) S' \tag{110a}$$

$$\frac{dS'}{dz} - \pi i s_g S' = \left(\frac{\pi i}{t_g}\right) T' \tag{110b}$$

The amplitudes T', S', T, and S are related as follows:

$$T = T' \exp(\pi i s_g z) \tag{111a}$$

$$S = S' \exp(-\pi i s_g z) \tag{111b}$$

We have shown in Sec. 1.1.4.4 of this Chapter that deformation can be

introduced in the diffraction equations by applying the deformable ion approximation, a defect being modeled by means of a vector field $\boldsymbol{R}(\boldsymbol{r})$, called its *displacement field*. It was also shown that the Fourier coefficients of the lattice potential then become functions of $\boldsymbol{r}$ and acquire a phase factor $V_g \exp(-i\alpha_g)$ with $\alpha_g = 2\pi \boldsymbol{g} \cdot \boldsymbol{R}$. Strictly speaking, the series given by Eq. (57) is then no longer a Fourier series, which is in agreement with the fact that the lattice potential is no longer periodic. It nevertheless turns out to be a good approximation, especially for small gradients of $\boldsymbol{R}(\boldsymbol{r})$. The presence of the defect can then be taken into account by substituting in the diffraction equations $V_g \to V_g \exp(-i\alpha_g)$. In view of the direct proportionality of $1/t_g$ and V_g [Eq. (41)], this implies that $1/t_g$ has to be replaced by $(1/t_g)\exp(-i\alpha_g)$. Equations (109a) and (109b) then take the form

$$\frac{dT}{dz} = \left(\frac{\pi i}{t_{-g}}\right) \exp(i\alpha_g) S \tag{112a}$$

$$\frac{dS}{dz} = 2\pi i s_g S + \left(\frac{\pi i}{t_g}\right) \exp(-i\alpha_g) T \tag{112b}$$

and Eqs. (110a) and (110b) become

$$\frac{dT'}{dz} + \pi i s_g T' = \left(\frac{\pi i}{t_{-g}}\right) \exp(i\alpha_g) S' \tag{113a}$$

$$\frac{dS'}{dz} - \pi i s_g S' = \left(\frac{\pi i}{t_g}\right) \exp(-i\alpha_g) T' \tag{113b}$$

Both systems of equations can be transformed so as to involve only the gradient of the displacement field.

Putting $T = T''$ and $S = S'' \exp(-i\alpha_g)$ the first system of equations [Eqs. (112a) and (112b)] is transformed into the set

$$\frac{dT''}{dz} = \left(\frac{\pi i}{t_{-g}}\right) S'' \tag{114a}$$

$$\frac{dS''}{dz} = 2\pi i \left(s_g + \frac{d\alpha'_g}{dz}\right) S'' + \left(\frac{\pi i}{t_g}\right) T'' \tag{114b}$$

with $\alpha'_g = \alpha_g / 2\pi$. Performing the substitution

$$T' = T''' \exp(\pi i \alpha'_g) \tag{115a}$$

$$S' = S''' \exp(-\pi i \alpha'_g) \tag{115b}$$

leads to the system

$$\frac{dT'''}{dz} + \pi i \left(s_g + \frac{d\alpha'_g}{dz}\right) T''' = (\pi i / t_{-g}) S''' \tag{116a}$$

$$\frac{dS'''}{dz} - \pi i \left(s_g + \frac{d\alpha'_g}{dz}\right) S''' = (\pi i / t_g) T''' \tag{116b}$$

The sets of Eqs. (113) and (116) are of the same form as the corresponding sets for the perfect crystal except for the s value. They suggest a simple interpretation: the presence of the defect causes the local s value, $s_{\text{eff}} \equiv s_g + \boldsymbol{g} \cdot d\boldsymbol{R}/dz$ to be, in general, different from its value in the perfect part of the foil. Locally, close to the defect, the Bragg condition is thus better or less well satisfied than in the rest of the foil, and hence the locally diffracted beam has a larger or smaller amplitude than in the perfect part.

If $\alpha = 2\pi \boldsymbol{g} \cdot \boldsymbol{R}_0 = \text{Constant}$, which is the case for a translation interface, with displacement vector $\boldsymbol{R}_0$, we have $d\alpha/dz = 0$ and the sets of Eqs. (114) and (116) reduce to those for a perfect crystal.

This is consistent with the fact that the two parts of the crystal on either side of the interface remain perfect. We have therefore to use the sets (113) or (112) to treat this problem, as we shall demonstrate below, rather than the sets (114) or (116).

If the interface is a domain boundary described by $\boldsymbol{R} = kz\boldsymbol{e}_\tau$ (see Sec. 1.1.1.4 of this Chapter) we have $\mathrm{d}\boldsymbol{R}/\mathrm{d}z = k\boldsymbol{e}_\tau$ and now the sets (114) and (116) can be used. They show that the crystal can be described as an assembly of two juxtaposed crystal parts separated by the interface and having different s values: s and $s + \Delta s$, with $\Delta s = k\boldsymbol{e}_\tau$.

For the displacement field of a dislocation, neither the factor $\exp(\mathrm{i}\alpha_{\boldsymbol{g}})$ nor the gradient $\mathrm{d}\boldsymbol{R}/\mathrm{d}z$ disappear (except for those $\boldsymbol{g}$ for which extinction occurs). The sets of Eqs. (114) and (116) as well as (112) and (113) are thus suitable descriptions in this case.

The anomalous absorption has not yet been introduced explicitly in the different sets of equations for faulted crystals. Phenomenologically this can easily be done by replacing $1/t_g$ by $(1/t_g + \mathrm{i}/\tau_g)$ or σ by $(\sigma_z + \mathrm{i}\sigma_\mathrm{i})$ in the sets of equations. Alternatively, it can be done by making the same substitution directly in the final results. We shall follow the latter approach for a discussion of the fringe patterns at inclined planar interfaces.

Matrix Formulation for the Amplitudes of Transmitted and Scattered Beams for Perfect Crystals

For a systematic discussion of the contrast at planar interfaces we make use of a matrix formulation for the amplitudes of the transmitted and scattered beams by a perfect crystal slab [24], which we now derive first.

Let T and S represent the transmitted and the scattered amplitudes for an incident wave with unit amplitude. The initial values at the entrance face of the slab are represented by the column vector $\binom{1}{0}$; at level z the amplitudes of transmitted and scattered beams are represented by the column vector $\binom{T}{S}$. From Sec. 1.1.1.5 of this Chapter we know that, ignoring a common irrelevant phase factor,

$$T \equiv T(z, s_g) = \cos(\pi\sigma_g z) - \mathrm{i}\left(\frac{s_g}{\sigma_g}\right)\sin(\pi\sigma_g z) \qquad (117)$$

$$S \equiv S(z, s_g) = \left(\frac{\mathrm{i}}{\sigma_g t_g}\right)\sin(\pi\sigma_g z) \qquad (118)$$

are the solutions of the sets of Eqs. (71) or (73) with initial values $\binom{1}{0}$. Anomalous absorption is taken into account by assuming $\sigma_g = \sigma_\mathrm{r} + \mathrm{i}\sigma_\mathrm{i}$ [Eq. (98)] with

$$\sigma_\mathrm{r} = \frac{[1 + (st_g)^2]^{1/2}}{t_g} \qquad (119a)$$

$$\sigma_\mathrm{i} = \frac{1}{(\sigma_\mathrm{r}\sigma_g\tau_g)} \qquad (119b)$$

In view of the linear character of the system of differential equations, for an incoming wave with an arbitrary amplitude we can write

$$\begin{pmatrix} T \\ S \end{pmatrix}_\mathrm{out} = \begin{pmatrix} A & C \\ B & D \end{pmatrix}\begin{pmatrix} T \\ S \end{pmatrix}_\mathrm{in} \qquad (120)$$

where the elements A, B, C, and D of the 2×2 matrix remain to be determined.

From

$$\begin{pmatrix} T(z,s_g) \\ S(z,s_g) \end{pmatrix} = \begin{pmatrix} A & C \\ B & D \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \equiv \begin{pmatrix} A \\ B \end{pmatrix} \tag{121}$$

we conclude that

$A = T$ and $B = S$.

We now make use of the symmetry of the system of Eqs. (71). We note that this system is mapped on itself by the substitution $T \to S$, $S \to T$, $s_g \to -s_g$, since $t_g = t_{-g}$ in a centrosymmetric crystal. This means that the solution for initial values $\binom{0}{1}$ is given by

$$\begin{pmatrix} S^- \\ T^- \end{pmatrix} = \begin{pmatrix} A & C \\ B & D \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \equiv \begin{pmatrix} C \\ D \end{pmatrix} \tag{122}$$

where the minus sign means: $T^{(-)} = T(z,-s_g)$ and $S^{(-)} = S(z,-s_g)$. We conclude that $C = S^{(-)}$ and $D = T^{(-)}$. The response matrix is thus completely defined for arbitrary initial values

$$\begin{pmatrix} T \\ S \end{pmatrix}_{\text{out}} = \begin{pmatrix} T & S^{(-)} \\ S & T^{(-)} \end{pmatrix} \begin{pmatrix} T \\ S \end{pmatrix}_{\text{in}} \tag{123}$$

We shall represent the response matrix M of a perfect crystal slab as

$$M(z,s) \equiv \begin{pmatrix} T & S^{(-)} \\ S & T^{(-)} \end{pmatrix} \tag{124}$$

Imagining a slab of perfect crystal with total thickness z_0 to be sliced in perfect slabs with thicknesses $z_1, z_2, \ldots, z_{n-1}, z_n$ such that $z_1 + z_2 + \cdots + z_n = z_0$ should clearly not influence the final result. We must therefore have

$$M(z_1 + z_2 + \cdots + z_n, s_g) = M(z_n, s_g) \cdot M(z_{n-1}, s_g) \ldots M(z_1, s_g) \tag{125}$$

This property of the response matrix can be verified by multiplying the matrices.

Equation (125) can formally be generalized to include also the subtraction of a lamella, that is

$$M(z_1 - z_2, s_g) = M(-z_2, s_g)\, M(z_1, s_g) \tag{126}$$

where

$$M(-z, s_g) = \begin{pmatrix} T(-z) & S^{(-)}(-z) \\ S(-z) & T^{(-)}(-z) \end{pmatrix} \tag{127}$$

1.1.1.10 Matrix Formulation for a Foil Containing a Translation Interface

A description of the diffraction effects associated with translation interfaces can be based on the set of Eqs. (113) [24]. The vector $\boldsymbol{R}_0$ describes the displacement of the exit part with respect to the entrance part of the foil; it determines the sign of $\alpha_g = 2\pi \boldsymbol{g} \cdot \boldsymbol{R}_0$. In the front part $\alpha_g = 0$, whereas in the exit part $\alpha_g \neq 0$. The translation interface at $z = z_1$ is assumed to be parallel to the foil surfaces. The total foil thickness is $z_1 + z_2 = z_0$. The front part being perfect and undisplaced; its response matrix is $M(z_1, s_g)$. Let the response matrix of the second part be represented by

$$\begin{pmatrix} U & X \\ V & Y \end{pmatrix} \tag{128}$$

where X, Y, U, and V must be determined from the set of Eqs. (113). We note that this set of equations reduces to that for a perfect undisplaced slab by means of the substitution $T' = T^{\text{s}}$, $S' = S^{\text{s}} \exp(-\mathrm{i}\alpha_g)$.

The solution of this set of equations is thus $T^s = T(z_2, s_g)$ and $S^s = S(z_2, s_g)$ since the front and exit parts have the same orientation. For the original set the solution is thus

$$T' = T(z_2, s_g) \tag{129a}$$

and

$$S' = S(z_2, s_g)\exp(-i\alpha_g) \tag{129b}$$

that is

$$\begin{pmatrix} T(z, s_g) \\ S(z, s_g)\exp(-i\alpha_g) \end{pmatrix} = \begin{pmatrix} U & X \\ V & Y \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \equiv \begin{pmatrix} U \\ V \end{pmatrix} \tag{130}$$

and hence

$$U = T(z_2, s_g) \tag{131a}$$

and

$$V = S(z_2, s_g)\exp(-i\alpha_g) \tag{131b}$$

We note that the system (113) is mapped onto itself by the substitution $T' \to S'$, $S' \to T'$, $s_g \to -s_g$, and $\alpha_g \to -\alpha_{-g}$. The solutions of this new set then also remain the same as those of the original set, except that the interchange of S' and T' has caused the initial values to become $\binom{0}{1}$ and that $s_g \to -s_g$. We thus find

$$\begin{pmatrix} S(z_2, -s_g)\exp(i\alpha_g) \\ T(z_2, -s_g) \end{pmatrix} = \begin{pmatrix} U & X \\ V & Y \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \equiv \begin{pmatrix} X \\ Y \end{pmatrix} \tag{132}$$

and

$$X = S(z_2, -s_g)\exp(i\alpha_g) \tag{133a}$$

and

$$Y = T(z_2, -s_g) \tag{133b}$$

Thus, in a more concise notation, the response matrix of the exit part is

$$\begin{pmatrix} T_2 & S_2^{(-)}\exp(i\alpha_g) \\ S_2\exp(-i\alpha_g) & T_2^{(-)} \end{pmatrix} \tag{134}$$

The response matrix of the faulted slab can thus be formulated as

$$\begin{pmatrix} T \\ S \end{pmatrix} = \begin{pmatrix} T_2 & S_2^{(-)}\exp(i\alpha_g) \\ S_2\exp(-i\alpha_g) & T_2^{(-)} \end{pmatrix} \times \begin{pmatrix} T_1 & S_1^{(-)} \\ S_1 & T_1^{(-)} \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{135}$$

Matrix (134) can conveniently be written as the product of three matrices:

$$\begin{pmatrix} T_2 & S_2^{(-)}\exp(i\alpha_g) \\ S_2\exp(-i\alpha_g) & T_2^{(-)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \exp(-i\alpha_g) \end{pmatrix}\begin{pmatrix} T_2 & S_2^{(-)} \\ S_2 & T_2^{(-)} \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & \exp(i\alpha_g) \end{pmatrix} \tag{136}$$

This suggests to introduce as a shorthand the *shift matrix*

$$\mathfrak{S}(\alpha_g) \equiv \begin{pmatrix} 1 & 0 \\ 0 & \exp(i\alpha_g) \end{pmatrix} \tag{137}$$

The final result for the response of the faulted slab can then be written as

$$\begin{pmatrix} T \\ S \end{pmatrix} = \mathfrak{S}(-\alpha_g)M_2\mathfrak{S}(\alpha_g)M_1\begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{138}$$

with

$$M_j \equiv M_j(z_j, s_j) = \begin{pmatrix} T_j & S_j^{(-)} \\ S_j & T_j^{(-)} \end{pmatrix} \tag{139}$$

The shift matrices have the property

$$\begin{pmatrix} 1 & 0 \\ 0 & \exp(i\alpha_1) \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \exp(i\alpha_2) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \exp[i(\alpha_1+\alpha_2)] \end{pmatrix} \quad (140)$$

that is $\mathfrak{S}(\alpha_1)\,\mathfrak{S}(\alpha_2) = \mathfrak{S}(\alpha_1+\alpha_2)$; they commute.

The result can be generalized directly to a succession of overlapping translation interfaces characterized by phase angles α_j, all referred to the front slab:

$$\begin{pmatrix} T \\ S \end{pmatrix} = \ldots \mathfrak{S}(-\alpha_2')M_3\mathfrak{S}(\alpha_2')\mathfrak{S}(-\alpha_1')M_2 \times \mathfrak{S}(\alpha_1')M_1\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (141)$$

Introducing the phase angles $\alpha_j = \alpha_j' - \alpha_{j-1}'$, which now describe the *relative* displacements of successive lamellae, the rear part being displaced with respect to the front one, we obtain

$$\begin{pmatrix} T \\ S \end{pmatrix} = \ldots M_3\mathfrak{S}(\alpha_2)M_2\mathfrak{S}(\alpha_1)M_1\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (142)$$

1.1.1.11 Matrix Formulation for a Foil Containing a Domain Boundary [20]

It is possible to generalize further by assuming that in successive lamellae the s values may be different, as in the case of domain boundaries.

For a pure domain boundary the transmitted and scattered amplitudes are given by

$$\begin{pmatrix} T \\ S \end{pmatrix} = M(z_2, s_{g,2})M(z_1, s_{g,1})\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (143)$$

where now $s_{g,1} \neq s_{g,2}$.

Equation (142) also describes the most general case of overlapping mixed boundaries, that is boundaries containing a translation component as well as exhibiting a difference in deviation parameters, provided the s values in the successive matrices, M_j, are assumed to be different.

1.1.1.12 Matrix Formulation for a Crystal Containing a Non-reflecting Part: the Vacuum Matrix

A foil may contain lamellae which are very far from any reflecting orientation under the diffraction conditions prevailing in the rest of the foil; except for absorption they behave as if no material was present in that lamellae. This is, for example, the case for a microtwin lamella in a face-centered structure if an uncommon reflection is excited in the matrix. It also applies to a cavity and to a precipitate lamella with a lattice different from that of the reflecting matrix. Even though such parts of the foil do not contribute to the diffraction phenomena, their presence influences the relative phases of the waves diffracted by the foil parts in front and behind these inactive lamellae. This can be accounted for by including the appropriate matrix in the matrix product, describing such a non-diffracting part [24].

In a nonreflecting part the extinction distance, t_g, is infinite and the system of

Eqs. (113) thus reduces to

$$\frac{dT}{dz} + \pi i s_g T = 0 \tag{144a}$$

and

$$\frac{dS}{dz} - \pi i s_g S = 0 \tag{144b}$$

which integrate to

$$T = T_0 \exp(-\pi i s_g z) \tag{145a}$$

and

$$S = S_0 \exp(\pi i s_g z) \tag{145b}$$

where T_0 and S_0 are the amplitudes at the entrance face of the non-reflecting 'vacuum' lamella. One can thus write

$$\begin{pmatrix} T \\ S \end{pmatrix}_{\text{out}} = \begin{pmatrix} \exp(-\pi i s_g z) & 0 \\ 0 & \exp(\pi i s_g z) \end{pmatrix} \times \begin{pmatrix} T \\ S \end{pmatrix}_{\text{in}} \tag{146}$$

The 'vacuum' matrix is thus

$$V(z, s_g) = \begin{pmatrix} \exp(-\pi i s_g z) & 0 \\ 0 & \exp(\pi i s_g z) \end{pmatrix} \tag{147}$$

where z is the thickness of the nonreflecting part as measured along the beam path and s_g is the deviation parameter of the crystal part *preceding* the 'vacuum' lamella.

1.1.1.13 Fringe Profiles at Planar Interfaces [25]

General Formulae

Multiplication of the matrices in Eq. (138) results in the following expressions for T and S:

$$T = T_1 T_2 + S_1 S_2^{(-)} \exp(i\alpha) \tag{148a}$$

and

$$S = T_1 S_2 \exp(-i\alpha) + S_1 T_2^{(-)} \tag{148b}$$

The minus signs in $S_2^{(-)}$ and $T_2^{(-)}$ indicate that the expressions S_2 and T_2 have to be modified by changing s to $-s$. Equations (148) express the fact that the transmitted amplitude results from the interference between the doubly transmitted beam $T_1 T_2$ and the doubly scattered beam $S_1 S_2^{(-)} \exp(i\alpha)$. The minus sign in $S_2^{(-)}$ indicates that the scattering, by the second part of the foil of the beam which has already been scattered by the first part, takes place from the $-\boldsymbol{g}$ side of the lattice planes. This implies that the deviation parameter has to be changed from s to $-s$, since for this second scattering event the node G acts as the origin of reciprocal space (Fig. 29). This second scattering

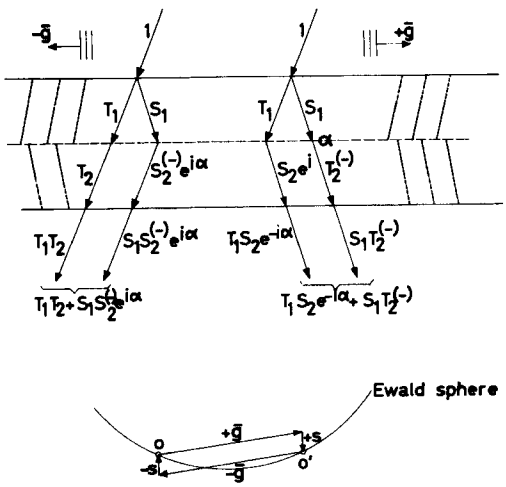


Figure 29. Transmitted and scattered amplitudes for a crystal containing a planar interface.

process is accompanied by a phase shift $\alpha = 2\pi \boldsymbol{g} \cdot \boldsymbol{R}$ due to the translation over $\boldsymbol{R}$ of the lattice planes in part II with respect to those in part I. This is taken into account by the phase factor $\exp(i\alpha)$. A similar interpretation can be given to the expression for S. The phase factor is now $\exp(-i\alpha)$ rather than $\exp(i\alpha)$ because the phase shifts due to the displacement of part II have opposite signs for $S_2^{(-)}$ and for S_2, since $\boldsymbol{g}$ has opposite sign for the two processes.

Introducing the explicit expressions for T_j and S_j ($j = 1$ or 2) given by Eqs. (117) and (118), in Eqs. (148a) and (148b), and noting that $\sigma = \sigma_r + i\sigma_i$ since anomalous absorption is taken into account, leads to explicit expressions for $I_T = TT^*$ and $I_S = SS^*$. It turns out that it is possible to cast these explicit expressions for I_T and I_S in a form which allows a detailed analytical discussion of the fringe profiles. Detailed calculations are given in [25] also for the general case, where the deviation parameters s_1 and s_2 as well as the extinction distances t_{g_1} and t_{g_2} in parts I and II, respectively, are assumed to be different. We do not give the details here of these straightforward but tedious calculations, but we will summarize the significant results.

The expressions for I_T and I_S can be written as sums of three terms

$$I_{T,S} = I_{T,S}^{(1)} + I_{T,S}^{(2)} + I_{T,S}^{(3)} \tag{149}$$

Translation Interfaces

We consider first pure translation interfaces [2]. Thus $s_1 = s_2 \equiv s$, and limiting ourselves to the case when $s = 0$, the expressions become

$$I_{T,S}^{(1)} = \tfrac{1}{2}\cos^2\left(\frac{\alpha}{2}\right) \times [\cosh(2\pi\sigma_i z_0) \pm \cos(2\pi\sigma_r z_0)] \tag{150}$$

$$I_{T,S}^{(2)} = \tfrac{1}{2}\sin^2\left(\frac{\alpha}{2}\right) \times [\cosh(4\pi\sigma_i u) \pm \cos(4\pi\sigma_r u)] \tag{151}$$

$$I_{T,S}^{(3)} = \tfrac{1}{2}\sin\alpha[\sin(2\pi\sigma_r z_1)\sinh(2\pi\sigma_i z_2) \pm \sin(2\pi\sigma_r z_2)\sinh(2\pi\sigma_i z_1)] \tag{152}$$

where the upper sign corresponds to I_T and the lower sign to I_S. The total thickness is $z_0 = z_1 + z_2$, where z_1 is the thickness of the front part and z_2 that of the rear part. Furthermore, we have $u = \frac{1}{2}(z_1 - z_2)$, that is u is the distance of the interface from the midplane of the foil. Along a planar interface intersecting the foil surfaces, as is often the case, z_1 and z_2 vary along the foil in such a way that $z_1 + z_2$ remains constant and equal to z_0. In the projected area of the interface fringes will be formed which, according to the column approximation, can be considered as being due to the intersection of the depth variation of I_T (or I_S) with the inclined interface.

If $\alpha = n \cdot 2\pi$, where n is an integer, which is the case if there is *no* stacking fault, $\sin\alpha = 0$ and $\sin(\alpha/2) = 0$. The only remaining term is then $I_{T,S}^{(1)}$, which as a result must represent the contribution due to the perfect crystal. This term only depends on the total thickness z_0; it describes a background onto which the fringes represented by the other terms are superimposed.

The second term, $I_{T,S}^{(2)}$, depends on u and not on z_1 and z_2 separately. It represents a function which is periodic in u, with a depth period $1/2\sigma_r$. The center of the

pattern, at $u = 0$, exhibits an extremum; it is a minimum for $I_S^{(2)}$ and a maximum for $I_T^{(2)}$. This pattern consists of fringes which are parallel to the central line of the pattern. We shall see that the amplitude of the fringes is only large enough to be visible in the central part of the pattern.

The dominant features of the pattern, in sufficiently thick foils, are described by $I_{T,S}^{(3)}$. Where the interface is close to the entrance face z_1 is small and $z_2 \simeq z_0$, and the factor $\sinh(2\pi\sigma_i z_2)$ is then large. The term $\frac{1}{2}\sin\alpha\sinh(2\pi\sigma_i z_2)\sin(2\pi\sigma_r z_1)$ represents a damped sinusoid with a depth period $1/\sigma_r$. This term disappears at the rear surface where $z_2 = 0$. If $\sin\alpha > 0$, the first extremum is a maximum; as a result the first fringe will be bright at the entrance face. For $\sin\alpha < 0$, the first fringe will be dark.

Where the interface is close to the exit face $z_1 \simeq z_0$ and $z_2 \simeq 0$; the term $\pm\frac{1}{2}\sin\alpha\sinh(2\pi\sigma_i z_1)\sin(2\pi\sigma_r z_2)$ is now dominant. It again represents a damped sinusoid. The first extremum, which now refers to the last fringe, is either a maximum or a minimum, depending on the sign of $\sin\alpha$. Note that the extremum is different for I_T and I_S since in the first case the plus sign applies, whereas for I_S the minus sign applies.

Figure 30 and Table 1 summarize this discussion. Note that these results, in particular those pertaining to the nature of the edge fringes, imply that anomalous absorption must be sufficiently large to make sure that the dominant behavior is described by the term $I_{T,S}^{(3)}$.

For a discussion of the behavior in the central part of the pattern the term $I_{T,S}^{(2)}$ may become important, since the two terms of $I_{T,S}^{(3)}$ compensate for certain thicknesses and $I_{T,S}^{(2)}$ may then become the dominant term.

For the important case $\alpha = \pm 2\pi/3$, which occurs in cubic-close-packed structures, the stacking fault fringes have the following properties, provided the foil is sufficiently thick so as to make anomalous absorption a dominant feature.

The bright field fringe pattern is symmetrical with respect to the line $u = 0$. This can be deduced quite generally from the implicit expressions [Eqs. (148)] for $s_1 = s_2 \equiv s$, which shows that $I_S = TT^*$ has the symmetry property

$$I_T(z_1, z_2, s, \alpha) = I_T(z_2, z_1, s, \alpha) \tag{153}$$

On the other hand, $I_S = SS^*$ has the property

$$I_S(z_1, z_2, s, \alpha) = I_S(z_2, z_1, -s, -\alpha) \tag{154}$$

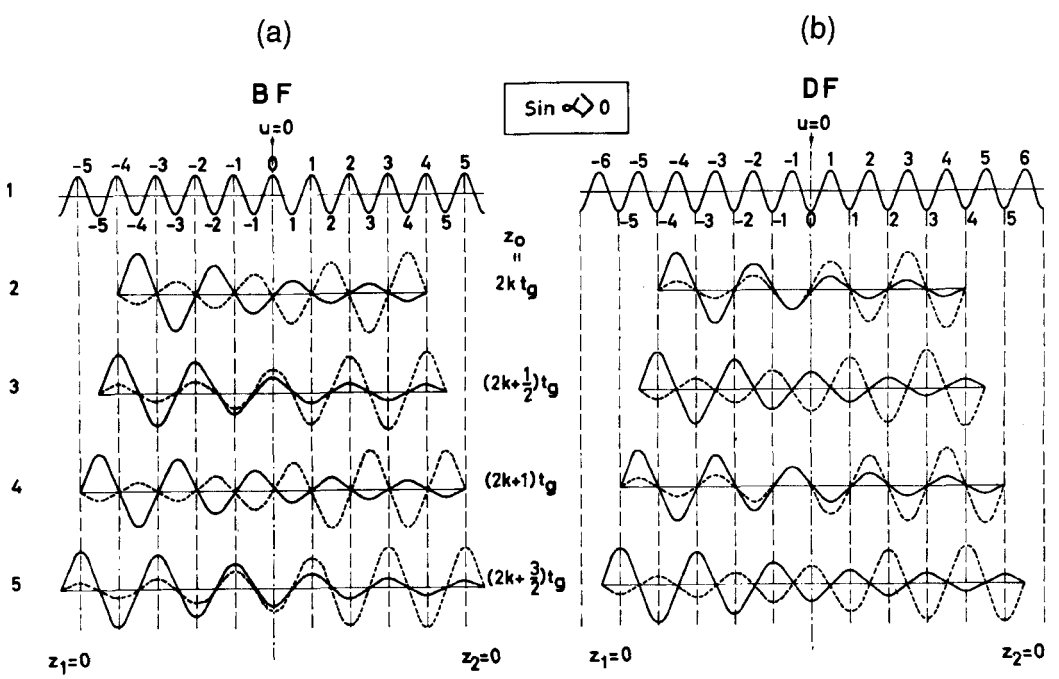
that is the dark field fringe pattern is antisymmetrical with respect to the foil center since interchanging z_1 and z_2 also requires changing the signs of s and α, which changes the nature of the edge fringes.

The fringes are parallel to the closest surface; as a result, new fringes caused by an increase in foil thickness are generated by fringe splitting close to the center of the pattern. This result can be understood by noting the relative shift, with increasing thickness, of the curves representing the two terms in $I_{T,S}^{(3)}$ in the central part of the foil where they overlap.

Close to the entrance face of the foil the fringe patterns are similar, but close to the exit part they are complementary. This property is generally true for diffraction contrast images when anomalous absorption is important; it is also true for dislocation images. The computed profiles shown in Fig. 31, which can be compared

Table 1. The properties of fringe profiles due to planar interfaces, showing the nature of the first and last edge fringes for different signs of $\sin\alpha$ and δ. A schematic profile and an observed pattern are given for the two types of fringe.

$\alpha = 2\pi(\boldsymbol{g}\cdot\boldsymbol{R})$					$\delta = s_1 t_{g_1} - s_2 t_{g_2}$				
	Bright field		Dark field			Bright field		Dark field	
	First	Last	First	Last		First	Last	First	Last
$\sin\alpha > 0$	B	B	B	D	$\delta > 0$	B	D	B	B
$\alpha \neq \pi$	–	–	–	–	–	–	–	–	–
$\sin\alpha < 0$	D	D	D	B	$\delta < 0$	D	B	D	D



with the observed fringes in a silicon wedge, allow verification of most of these properties.

It is clear from the foregoing discussion that the nature of the edge fringes depends on the sign of $\sin\alpha$. The case when $\alpha = \pi$ is singular, because now $\sin\alpha = 0$ and the term $I_{\mathrm{T,S}}^{(3)}$ (which is mostly dominant) is now absent, as is the background term $I_{\mathrm{T,S}}^{(1)}$ since $\cos(\alpha/2) = 0$. The complete fringe pattern for $s = 0$ is then represented by

$$I_{\mathrm{T,S}}^{(2)} = \tfrac{1}{2}[\cosh(4\pi\sigma_{\mathrm{i}}u) \pm \cos(4\pi\sigma_{\mathrm{r}}u)] \quad (155)$$

since $\sin(\alpha/2) = 1$.

This expression shows that the bright field and dark field images are complementary with respect to the nonperiodic background which is described by $\cosh(4\pi\sigma_{\mathrm{i}}u)$. This background exhibits a

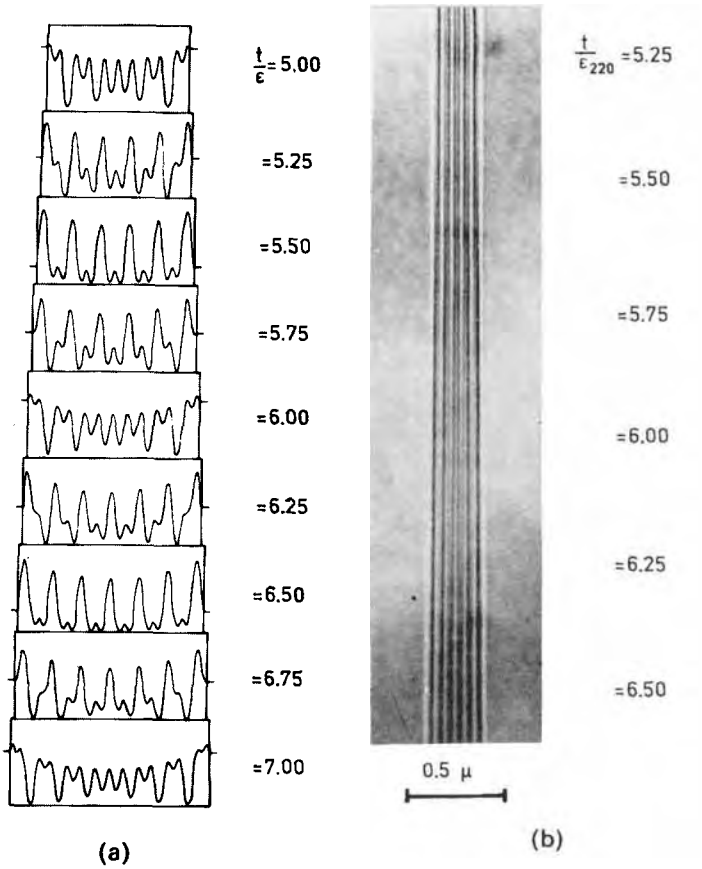


Figure 31. (a) Computed profiles for a stacking fault with increasing thickness of the specimen. (b) Fringe pattern due to a stacking fault in a wedge-shaped foil of silicon ($s = 0$; $g = 220$; $\alpha = 2\pi/3$; thickness $= 5t_g - 7t_g$.) (Courtesy of G. R. Booker.)

minimum for $u = 0$, that is it shows a minimum in the central part of the pattern. The fringes with a depth period $1/(2\sigma_r)$ are parallel to the central lines $u = 0$; they are superimposed on this background. As the thickness increases, new fringes are added at the surfaces.

Domain Boundary Fringes

We now consider interfaces separating two juxtaposed crystal parts in which the deviation parameters for homologous, simultaneously excited reflections are slightly different [25]. Such boundaries occur, for instance, in microstructures that result from phase transitions in which rotation symmetry elements are lost. Often the interface is a coherent twin with a small twinning vector. Under these conditions the simultaneously excited diffraction vectors $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ differ by $\Delta\boldsymbol{g} = \boldsymbol{g}_2 - \boldsymbol{g}_1$, where $\Delta\boldsymbol{g}$ is perpendicular to the coherent twin interface (Fig. 32). The difference in the deviation parameters, $\Delta s = s_2 - s_1$, is the projection of $\Delta\boldsymbol{g}$ along the normal to the foil plane. In the 'symmetrical' case, one has $s_2 = \frac{1}{2}\Delta s$ and $s_1 = -\frac{1}{2}\Delta s$, that is $s_1 = -s_2$.

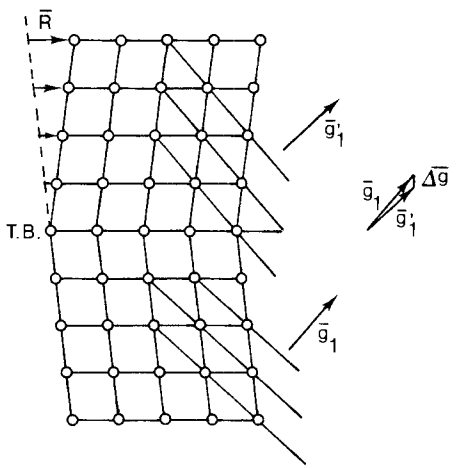


Figure 32. Schematic representation of a coherent domain boundary. TB, (Twin boundary).

The expressions for the transmitted and scattered amplitudes for a pure domain boundary (i.e., without a translation component) are given by

$$T = T_1 T_2 + S_1 S_2^- \tag{156a}$$

and

$$S = TS_2 + T_2^{(-)} S_1 \tag{156b}$$

of which the interpretation has been discussed in detail for the case of a stacking fault.

The general expressions for the intensities I_T and I_S, assuming the extinction distances in the two parts to be different, can again be written as the sum of three terms, and a discussion similar to the one given for translation interfaces is possible [25]. For sufficiently thick foils the behavior is again dominated by the terms $I^{(3)}_{T,S}$, which we will now discuss.

The general features of the fringe pattern are adequately exhibited by the 'symmetrical' case defined above. The terms $I^{(3)}_{T,S}$ are now given by [25]

$$\begin{aligned} w^4 I^{(3)}_{T,S} = &-\tfrac{1}{2}\delta\{\cos(2\pi\sigma_{r,1}z_1) \\ &\times \sinh[2(\pi\sigma_{i,2}z_2 \pm \varphi_2)] \\ &\mp \cos(2\pi\sigma_{r,2}z_2) \\ &\times \sinh[2(\pi\sigma_{i,1}z_1 \pm \varphi_1)]\} \end{aligned} \tag{157}$$

with

$$w^2 = 1 + (st_g)^2$$

$$t_g = \tfrac{1}{2}(t_{g_1} + t_{g_2})$$

$$\delta = s_1 t_{g,1} - s_2 t_{g,2}$$

$$2\varphi_j = \operatorname{argsinh}(st_{g,j})$$

and

$$st_g = \tfrac{1}{2}(s_1 t_{g_1} + st_{g_2})$$

The upper sign applies to I_T and the lower to I_S. The nature of the fringes is visibly determined by the parameter δ, which is assumed to be sufficiently small so that the same $\boldsymbol{g}$ vector is excited in both crystal parts.

Close to the front surface the first term of $I^{(3)}_{T,S}$ determines the behavior since $\sinh(2\pi\sigma_{i,2}z_2 \pm \varphi_2)$ is large for $z_2 \simeq z_0$ and $z_1 \simeq 0$. Close to the exit face the second term is dominant. The dependence of the nature of the edge fringes on the sign of δ is summarized in Table 1 and in Fig. 33.

The most striking and useful feature is the difference in the symmetry of the fringe patterns, due to translation interfaces. Whereas the bright field pattern for a translation interface is symmetrical with respect to the central line, it is roughly anti-symmetrical for a domain boundary, the edge fringes being opposite in nature. On the other hand, for the special case $s_1 = -s_2$ and $t_{g_1} = t_{g_2}$ the dark field image

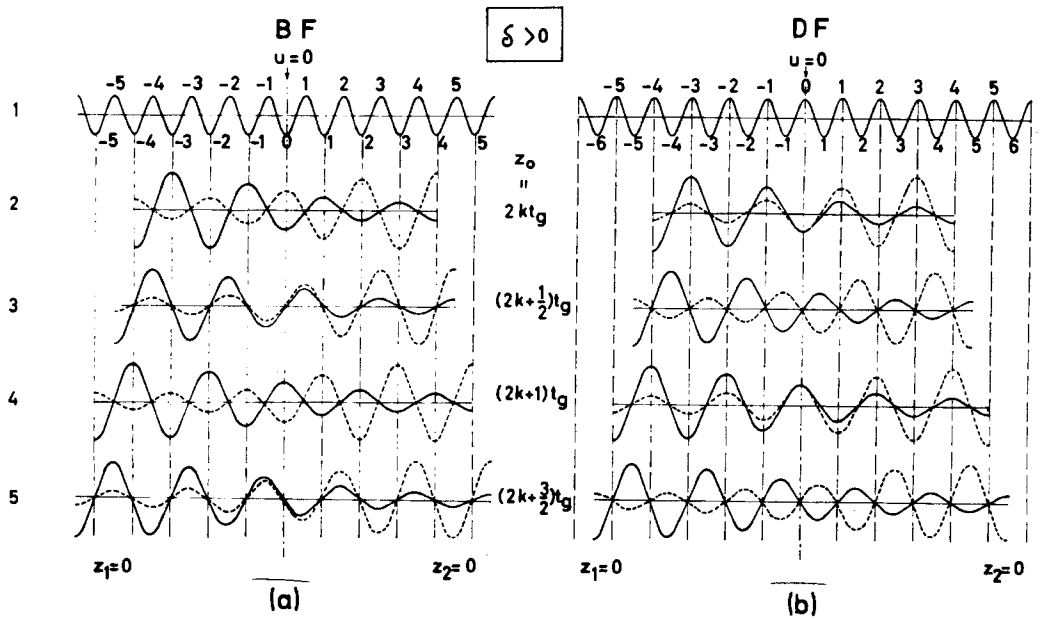


Figure 33. Schematic representation of the fringe profile due to a coherent domain boundary [25]: (a) I_T; (b) I_S. BF, bright field; DF, dark field. I_T (Transmitted Intensity), I_S (Scattered Intensity).

is symmetrical for domain boundary fringe patterns, but anti-symmetrical for translation interfaces. If t_{g_1} and t_{g_2} are significantly different, the depth periods close to the front and rear surfaces may be different. As for translation interfaces, the fringes are parallel to the closest surface; this is a consequence of anomalous absorption and it is therefore only true in sufficiently thick foils.

A characteristic feature of domain boundary images is that the domain contrast on either side of the interface (i.e., on either side of the fringe pattern) may be different, which is never the case for translation interfaces. However, for $s_1 = -s_2$ (i.e., in the symmetrical situation) the domain contrast is the same in both domains in the dark field image, but not in the bright field image. This is a consequence of the symmetry of the rocking curve for the scattered beam, on the one hand, and its asymmetry for the transmitted beam (the Borrmann effect), on the other.

Along certain interfaces there may be a phase shift as well as a difference in the orientation or length of the excited diffraction vector in the two crystal parts. The fringes produced along such interfaces have properties which are intermediate between those of pure α fringes and pure δ fringes [25].

Extinction Criteria

It is clear that no α fringes are produced if $\boldsymbol{g} \cdot \boldsymbol{R} =$ Integer. In Eqs. (150)–(152) for $I_{T,S}$, the terms $I_{T,S}^{(2)}$ and $I_{T,S}^{(3)}$ become zero. Only $I_{T,S}^{(1)}$ is different from zero; however, this term represents thickness fringes,

as it only depends on z_0. In fact it is easy to verify that for a perfect crystal Eqs. (151) and (107) are identical (for $s = 0$).

If an image is made using a diffraction vector which is common to the two crystal parts, that is if a diffraction spot belonging to the unsplit row or the unsplit plane is selected, the δ component of a mixed interface becomes inoperative and only a possible translation component may produce α fringes. It is also possible to eliminate selectively the translation component from the images of mixed boundaries. In this way it is, for example, possible to image the lattice relaxation along antiphase boundaries or stacking faults with a displacement vector $\boldsymbol{R}_0 + \boldsymbol{\varepsilon}$ by exciting only a systematic row of reflections $(\ldots -2\boldsymbol{g}, -\boldsymbol{g}, 0, +\boldsymbol{g}, +2\boldsymbol{g} \ldots)$ for which $\boldsymbol{g} \cdot \boldsymbol{R} =$ Integer. The presence of relaxation is then revealed by the occurrence of weak residual fringes, due to the additional displacement $\boldsymbol{\varepsilon}$, for which $\boldsymbol{g} \cdot \boldsymbol{\varepsilon} \neq$ Integer [26]. Using a number of different reflections for which $\boldsymbol{g} \cdot \boldsymbol{R}_0 =$ Integer but for which $\boldsymbol{g} \cdot \boldsymbol{\varepsilon} \neq$ Integer, one can obtain a fair idea of the direction, sense, and magnitude of $\boldsymbol{\varepsilon}$ from observations of the nature and contrast of the edge fringes in the residual fringe patterns.

1.1.1.14 Domain Fragmented Crystals: Microtextures

Many phase transformations lead to a decrease in space-group symmetry, the space group of the low-temperature phase being a subgroup of that of the high-temperature phase. As a result, a single crystal of the high-temperature phase usually becomes fragmented into domains after transformation into the low-temperature, low-symmetry phase. The structures within these domains are then related by symmetry operations lost during the transformation. The lost symmetry elements can be either *translations* or *rotations* [27]. The interfaces resulting from lost translation symmetry are translation interfaces (e.g., antiphase boundaries, stacking faults, discommensurations, and crystallographic shear planes). Lost rotation symmetry elements give rise to twins or domain boundaries. The use of the term *domain boundary* will be reserved for those cases where the lattices of the two domains are only slightly different. The reciprocal lattice nodes belonging to the two domains are then sufficiently close to each other to be excited simultaneously, albeit with different deviation parameters, and produce δ fringes.

At the other extreme, if the diffraction spots in a diffraction pattern made across the interface are sufficiently split so as to be able to make a dark field image in one of the components separately, we call the interface a *twin*. The image so obtained then exhibits wedge fringes in the selected domain. It is clear that the distinction between twins and domain boundaries is not very strictly defined in this way and intermediate situations are possible.

In some cases the lattices of the two domains separated by the interface are the same but the structures may be different. This is, for example, the case in noncentrosymmetrical crystals where the structures in the domains may be related by an inversion operation, or by a two-fold axis, the lattice being unperturbed by the interface. The domain structure in α-quartz provides an example of the latter type. The high-temperature β form of

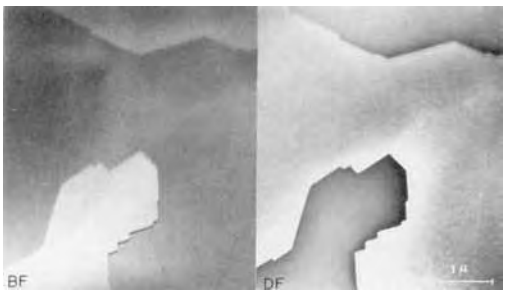


Figure 34. Dauphiné twin domains in α-quartz revealed by domain contrast [28–30]. BF, bright field; DF, dark field.

quartz has point-group symmetry 6 2 2, whereas the low-temperature α form belongs to the point group 3 2, that is the six-fold axis of the β phase becomes a three-fold axis in the α phase. On cooling to below the $\beta \rightarrow \alpha$ transition temperature (about 573°C), the β phase breaks up into Dauphiné twins, α_1 and α_2, of the α phase. The structures of α_1 and α_2 are related by the lost 180° rotation about the three-fold axis, whereas the lattice of α_1 and α_2 remains common (Fig. 34) [28–30].

Inversion boundaries occur in many non-centrosymmetric crystals and, for example, in the cubic χ-phase alloy Fe-Cu-Mo-Ti (Fig. 35) [31].

Domain textures can conveniently be studied by a combination of diffraction and imaging techniques exploiting different diffraction contrast phenomena.

1.1.1.15 Diffraction Patterns of Domain Textures

Domain textures produce a composite diffraction pattern which is the superposition of the diffraction patterns of the separate domains. This usually affects the geometry of the diffraction pattern by the occurrence of split spots; in some cases only the intensities are changed compared to those of a monodomain pattern.

Where differently oriented domains overlap when viewed along the zone axis, double diffraction may occur. This may sometimes complicate the interpretation by simulating a diffraction pattern with lower translation symmetry than that of

Figure 35. Inversion domains in the χ phase of Fe-Cr-Mo-Ti revealed by domain contrast [31].

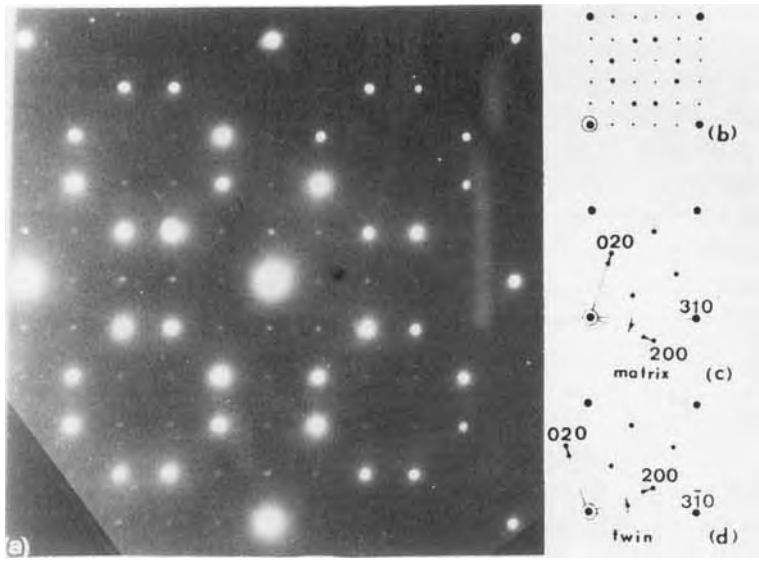


Figure 36. Diffraction pattern of Ni_4Mo exhibiting weak double-diffraction spots. (a) observed pattern; (b) Schematic: small dots represent double diffraction spots; (c) one orientation variant; (d) second orientation variant.

the separate domains, as a result of the double-diffraction spots (Fig. 36).

If several orientation variants are present in the selected area the diffraction pattern may become quite complicated and difficult to 'unscramble'. The interpretation can be simplified by first making monodomain diffraction patterns of the domains on both sides of the interface and subsequently from an area across the interface separating the two domains. However, this is only possible if the domains are sufficiently large.

The diffraction patterns across twins have characteristic features which allow one to determine the twinning elements. The reciprocal space of a reflection twin is represented in Fig. 37a; it exhibits a central row of unsplit nodes, perpendicular to the mirror plane in real space. This is a general feature of the relationship between direct and reciprocal space. A common lattice plane in real space (the coherent mirror plane) is represented in reciprocal space as a common lattice row perpendicular to the mirror plane. A common lattice row in direct space, as is the case for the lattice row along a 180° rotation twin axis, is represented in reciprocal space as a common reciprocal lattice plane perpendicular to the twinning axis. All other spots are split (Fig. 37b) [32].

In the case of a reflection twin the spot splitting is parallel to the unsplit row and its magnitude is proportional to the distance from the unsplit row. The magnitude of the spot splitting is a direct measure for the twinning vector. For a 180° rotation twin, the spots are all split along a direction parallel to the unsplit plane; the magnitude of splitting is proportional to the distance from the unsplit plane.

It is not always obvious how to distinguish between the two cases, since many sections of reciprocal space will look very similar. Tilting experiments exploring the relevant parts of reciprocal space are required in order to differentiate between the two cases. An example of the type of

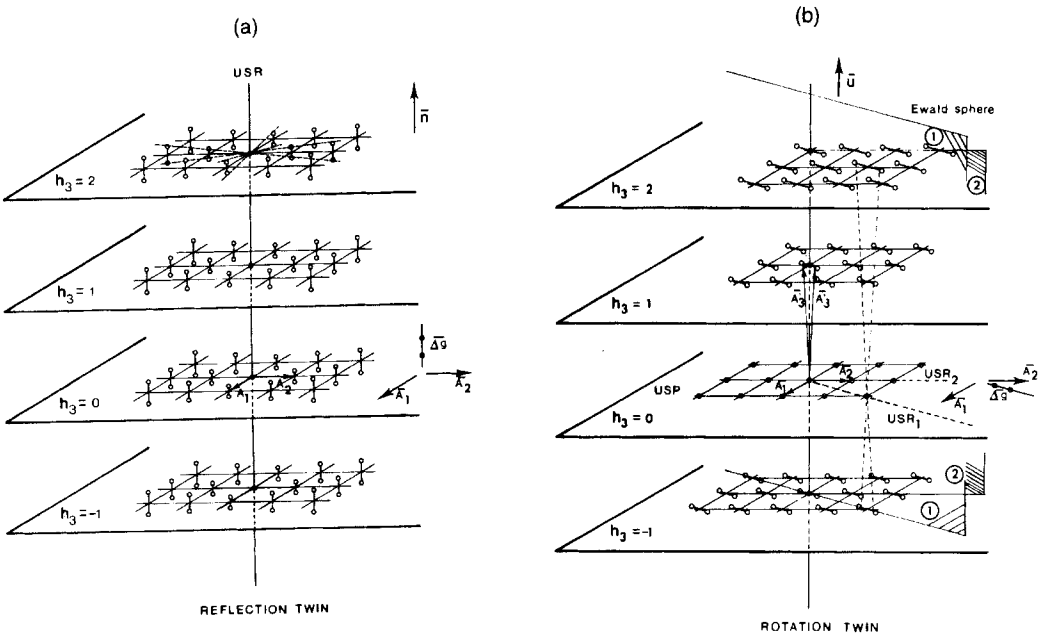


Figure 37. Reciprocal lattice of (a) a reflection twin and (b) a 180° rotation twin [32]. USR, (unsplit row); USP (unsplit plane).

experiment to be performed is shown in Fig. 38 [33].

The presence of higher order symmetry elements relating the structures in the different domains is reflected in the symmetry of the diffraction pattern. Figure 39 shows, for instance, the presence of three orthorhombic orientation variants related by 120° rotations along the zone axis. Such microstructures can usually be analyzed in terms of reflection or 180° rotation twins, by considering pairs of domains.

The distinction between a diffraction pattern produced by a quasicrystal along a noncrystallographic zone (e.g., a five-fold or ten-fold symmetry axis) and a diffraction pattern due to multiply-twinned 'classical' crystals is not always obvious, and has given rise to much debate.

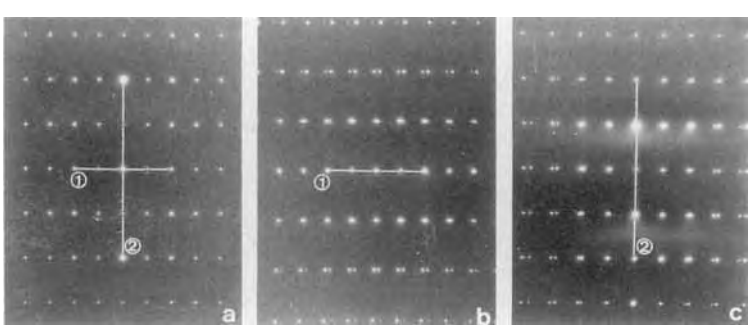


Figure 38. Tilting experiment in $MoTe_2$ allowing the presence of 180° rotation twins to be shown [33]. 1 and 2 are the tilt axis.

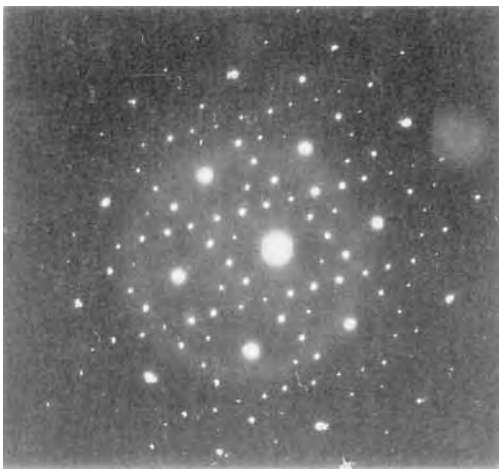


Figure 39. Composite diffraction pattern of a foil containing three orthorhombic orientation variants of Ni_3Mo related by 120° rotations. (Courtesy of G. Van Tendeloo.)

The presence of domains which are built on a common lattice is not reflected in the geometry of the diffraction pattern since it causes neither spot splitting nor additional spots compared with a mono-domain pattern. The relative intensities of the spots are affected, but this is not easily detected in electron diffraction patterns. Imaging techniques are of considerable help in the study of such textures.

1.1.1.16 Imaging of Microtextures

Microtextures can be imaged either by means of domain contrast, by interface contrast, or by both.

Domain Contrast of Orientation Variants

Domain contrast usually finds its origin in a small difference in the deviation parameters in adjacent domains leading to a significant difference in brightness in either the bright field image or the dark field image, made in a split reflection. More pronounced contrast arises if the dark field image is made in one of the components of a split reflection. However, this is only possible if the spot splitting is large enough.

The difference in brightness in the bright field image can be understood with reference to the asymmetric rocking curve for I_T. In the vicinity of $s = 0$ the s dependence of I_T is quite steep and a small difference in deviation parameter leads to a pronounced difference in transmitted intensity. Optimum domain contrast is thus obtained if the average deviation parameter is close to $s = 0$.

In the dark field image optimum contrast is achieved if a single spot can be isolated. If this is not the case the symmetry of the rocking curve for I_S shows that now the optimum contrast is obtained for an average s which is different from zero.

Domain contrast can also arise because the moduli of the structural factors, and hence the extinction distances, are different in adjacent domains. This is, for instance, the case for Dauphiné twins in quartz [28–30]. The lattices of α_1 and α_2 coincide, no spot splitting occurs and the above-mentioned contrast phenomena are inoperative. However, a number of coinciding reflections have structure amplitudes of different magnitudes. Dark field images made in such reflections will give rise to domain contrast, often called *structure factor contrast* (see Fig. 34).

It is clear that translation variants cannot give rise to domain contrast since the lattices, as well as the structures, are strictly parallel in the two domains.

Interface Contrast

It is also possible to image the interfaces rather than the domains. This is the only possibility for translation interfaces. For orientation variants domain contrast and interface contrast are often produced simultaneously.

The interfaces separating translation variants such as out-of-phase boundaries, crystallographic shear planes and stacking faults are imaged as α-type fringes in reflections for which $\boldsymbol{g} \cdot \boldsymbol{R}_0 \neq$ Integer. This is also the case for interfaces separating structural variants built on a common lattice, but having different structure amplitudes. We have seen above that domain contrast arises as a result of structure factor contrast when the moduli of the structure factors are different. However, it often happens that the structure factors have the same modulus in the two domains, but have a different phase. This is the case for certain reflections in domain fragmented α-quartz. A dark field image in such a reflection will not exhibit domain contrast but will reveal the interfaces as α-type fringe patterns.

The phase angle, α, can be deduced as follows. The structure factors for the structures in the two domains is written with respect to a common origin. The two structure factors are then related as follows:

$$F_{\boldsymbol{H}}^{(2)} = F_{\boldsymbol{H}}^{(1)} \exp(\mathrm{i}\alpha_{\boldsymbol{H}}) \qquad (158)$$

where α is the phase angle characterizing the fringe pattern, $F_{\boldsymbol{H}}^{(2)}$ and $F_{\boldsymbol{H}}^{(1)}$ are the structure factors with indices $\boldsymbol{H}$ in the exit and front part, respectively.

Inversion boundaries revealed by means of interface contrast in the χ phase of the alloy Fe-Cr-Mo-Ti are visible in Fig. 40 [31, 34–36].

Figure 40. Inversion domains in the χ phase of Fe-Cr-Mo-Ti, as revealed by interface contrast [34].

A particular type of interface contrast arises in dark field images made in double-diffraction spots caused by overlapping orientation domains. The projected interfacial region will now appear bright since double diffraction is only produced in the regions of overlap along the interfaces (Fig. 41).

The contrast at discommensuration walls is somewhat similar to that at antiphase boundaries. In the latter case dark field imaging in superstructure spots is used, whereas discommensuration walls are best imaged in dark field images, selecting incommensurate diffraction spots due to the modulation.

Inversion Boundaries

The first observations of inversion boundaries were made on the cubic noncentrosymmetric χ phase in the alloy system Fe-Cr-Mo-Ti [34–36] (see Fig. 35). The contrast at this type of boundary requires some specific discussion. It has been found experimentally that under the appropriate diffraction conditions the domain structure can be revealed by domain contrast as well as by interface contrast. Inversion

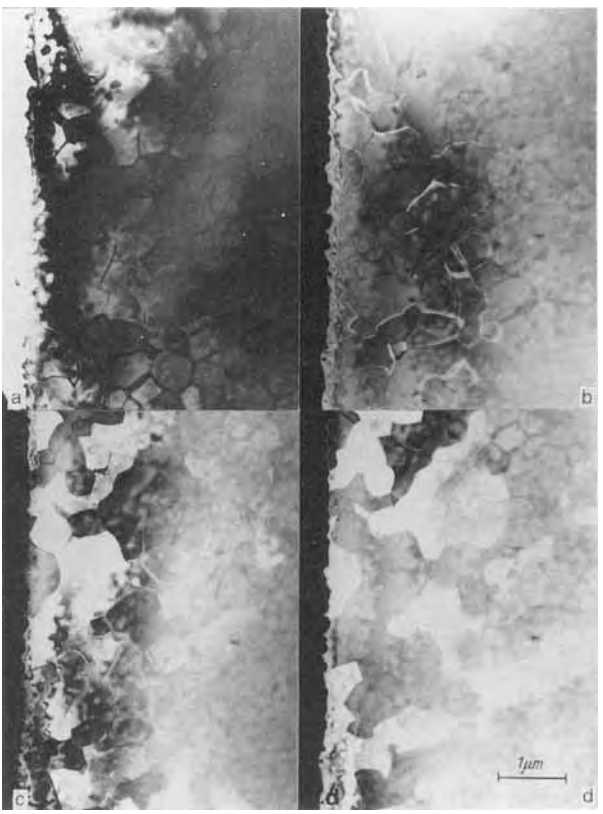


Figure 41. Interfaces in Ni_4Mo revealed by different imaging modes: (a) BF image; (b) in double-diffraction spot; (c) DF image in 020 of one variant; (d) DF image in 020 of second variant.

domains have a common lattice and hence there is no splitting. The structures are related by an inversion operation, that is the reflections $\boldsymbol{H}$ in one domain and $-\boldsymbol{H}$ in the other domain are always excited simultaneously and to the same extent. The moduli of the structure factors of simultaneously excited reflections $\boldsymbol{H}$ and $-\boldsymbol{H}$ are always the same according to Friedel's law $I_{\boldsymbol{H}} = I_{-\boldsymbol{H}}$. The phases $\alpha_{\boldsymbol{H}}$ and $\alpha_{-\boldsymbol{H}}$ are different for most reflections since the structure is noncentrosymmetric. For a noncentrosymmetric crystal the phases associated with the Fourier coefficients of the imaginary part of the lattice potential need not be equal to those associated with the Fourier coefficients of the real part.

Serneels et al. [34] have shown that domain contrast arises as a result of the violation of Friedel's law in dark field images in non-centrosymmetric crystals under multiple beam conditions, along a zone which does not produce centrosymmetry in projection. This means that the zone axis cannot be a symmetry axis of even order. For example, in the χ phase domain contrast is produced when the zone axis is along the three-fold rotation axis.

Interface contrast arises as a result of the difference in phase of the structure factors associated with the different domains. The interfaces are imaged as α fringes, the lattices being parallel in the

domains. The difference in phase between the Fourier coefficients related to real and imaginary parts of the lattice potential leads to weak interface contrast, even under two-beam conditions [34].

1.1.1.17 Dislocation Contrast

Intuitive Considerations

Dislocations are usually visible as dark lines in two-beam diffraction contrast images, made with small values of the deviation parameter. When applying the weak beam method, that is for large values of s, the dislocations appear as bright lines on a darker background.

The two-beam image formation at dislocations can easily be understood on intuitive grounds [18–20]. The foil represented in Fig. 42 contains an edge dislocation in E. The lattice planes used for imaging are indicated schematically. Due to the presence of the dislocation the lattice planes in the vicinity of E are slightly curved and inclined in opposite

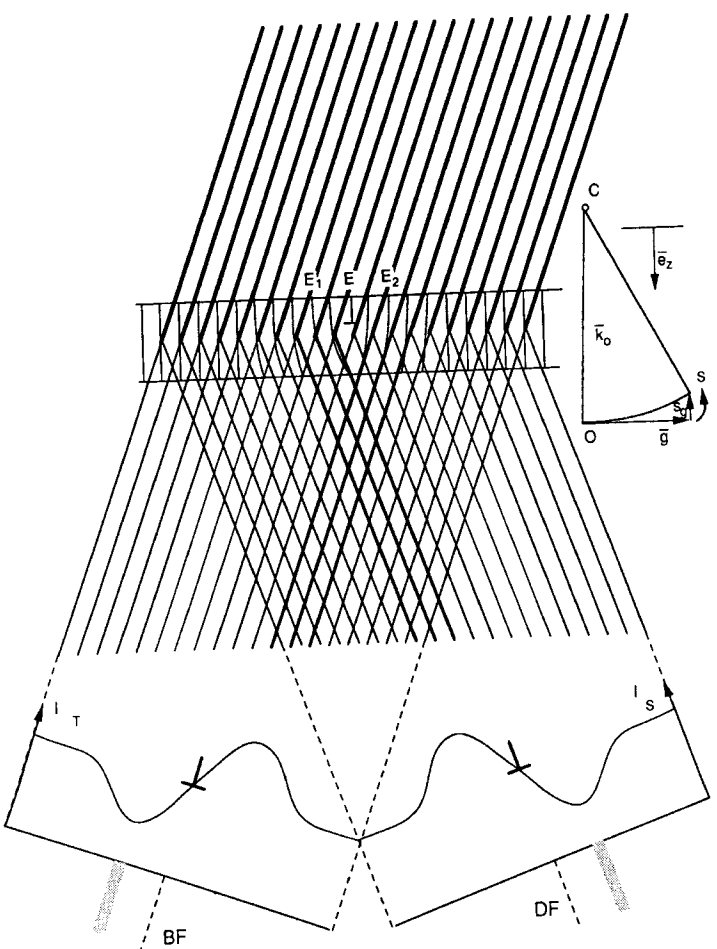


Figure 42. Intuitive model for the origin of contrast at an edge dislocation E. The thickness of the lines is a measure of the intensity of the electron beams. BF, bright field; DF, dark field.

senses to the left and right of E. As the specimen is a thin foil, the Bragg condition is relaxed; the reciprocal lattice nodes have become 'relrods'. We can therefore assume that diffraction occurs, even though the Bragg condition is only approximately satisfied with $s < 0$ in the part of the foil which is not affected by the presence of the dislocations. On the left of the dislocation, at E_1, the rotation of the lattice planes is then such that, locally, the Bragg condition is better satisfied, that is $|s|$ is smaller, and hence the diffracted beam will be more intense than in the perfect parts of the foil. On the right of the dislocation in E_2, the lattice rotation is in the opposite sense and hence the diffracted beam locally will be weaker than in the perfect part of the foil. The relative intensities of the diffracted beams are indicated schematically by lines of different widths in Fig. 42. Since no electrons are lost, the transmitted beam will be depleted where the scattered beam is enhanced.

Selecting the diffracted beam by means of an aperture and magnifying the corresponding diffraction spot will produce a map of the intensity distribution in this beam. This map will reveal a lack of intensity (i.e., a dark line) to the right of the dislocations in E_2 and an excess of intensity over the background in E_1. The dislocation will thus be imaged as a bright–dark line pair. This image is called a *dark field image*.

When selecting the transmitted beam a similar intensity map can be produced by magnifying the intensity distribution in the direct beam. Such an image is called a *bright field image*; in this approximation it is complementary to the dark field image, bright and dark lines being interchanged. Bright and dark field images are, in fact, only small portions of strongly magnified diffraction spots, the intensity distribution being the image.

The possibility of forming such images is a consequence of the 'local' character of electron diffraction. Electrons only sense a narrow column of material because the Bragg angles are small, electron diffraction is strongly peaked forward, and the foil is thin. The columns form, in a sense, the 'pixels' of the image. The assumption that electrons travel in narrow columns is the basis of the 'column approximation' introduced above.

The same type of reasoning can be used to demonstrate that screw dislocations produce a line image. As a consequence of the presence of the screw dislocation, the families of lattice planes intersecting the dislocation line are transformed into helical surfaces. To the left and right of the dislocation the lattice planes are slightly inclined in opposite senses, and hence the local diffraction conditions are different to the left and the right. Again a bright–dark line is produced (Fig. 43).

In both cases, the dark line image is *not* produced at the dislocation core, but in a slightly displaced position called the *image side*. Changing the diffraction conditions so as to make $s > 0$ in the foil part which is far away from the dislocations, changes the image side, as can be demonstrated by the same reasoning used above. Also, changing $\boldsymbol{g}$ to $-\boldsymbol{g}$ changes the image side, since now reflection takes place from the other side of the lattice planes. Finally, changing the sign of the Burgers vector changes the sense of inclination of the lattice planes on a given side of the dislocation, and hence also changes the image side. Summarizing, we can say that the image side depends on the sign of $p = (\boldsymbol{g} \cdot \boldsymbol{b})s$.

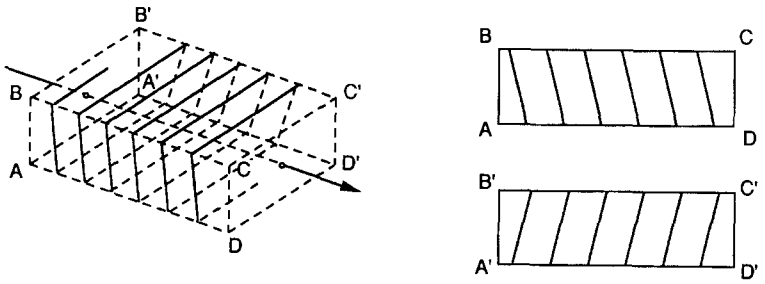


Figure 43. Geometry of lattice planes in the vicinity of a screw dislocation, leading to image contrast.

This rule becomes undetermined if $\boldsymbol{g} \cdot \boldsymbol{b} = 0$. The relation $\boldsymbol{g} \cdot \boldsymbol{b} = 0$ is, in fact, the criterion for the absence of contrast. It expresses the fact that no image is produced when diffraction occurs by the lattice planes which are left undeformed by the presence of the dislocation. To a first approximation, all displacements around a dislocation are parallel to the Burgers vector and are thus parallel to the lattice planes for which $\boldsymbol{g} \cdot \boldsymbol{b} = 0$. This *extinction criterion* is strictly valid for screw dislocations in an elastically isotropic medium, for which all displacements are parallel to $\boldsymbol{b}$, but it is only a first approximation for edge dislocations. Deviations occur even for screws in strongly anisotropic media, the reason being that the actual extinction criterion is $\boldsymbol{g} \cdot \boldsymbol{R} = 0$. The displacement field of an edge dislocation contains a component perpendicular to the glide plane, which causes some residual contrast even if $\boldsymbol{g} \cdot \boldsymbol{b} = 0$, as we shall discuss below.

Some contrast may also result, even though $\boldsymbol{g} \cdot \boldsymbol{b} = 0$, from the fact that the specimen is a thin foil. The presence of dislocations in a thin foil modifies the displacement field as a result of surface relaxation effects and this may produce contrast. For instance, a pure screw dislocation parallel to the incident beam and perpendicular to the foil surfaces is not expected to produce any contrast since $\boldsymbol{g} \cdot \boldsymbol{b} = 0$ for all active $\boldsymbol{g}$ vectors. However, such dislocations produce a dark–bright dot contrast which was attributed to the lattice twist. It was shown by Eshelby and Stroh [35] that, close to the emergence point of a screw dislocation in the foil surfaces, significant elastic relaxation takes place which transforms the lattice planes parallel to the dislocation line into helical surfaces, the sense of the helical twist being determined by the sign of the screw dislocation. This helical twist produces a bright–dark dot pair because on one side of the emergence point the lattice planes are tilted *into* the Bragg condition and on the other side they are tilted *out of* the Bragg condition. The line joining the bright–dark dot pair is perpendicular to $\boldsymbol{g}$ (Fig. 44a). Depending on the sense of the helical twist (i.e., on the sign of the screw dislocation), the dot pair is bright–dark or dark–bright. The sign of the screw dislocation can thus be determined from such images. Images of this type observed in a platinum foil are shown in Fig. 44b.

An edge dislocation parallel to the foil surfaces and with its glide plane also parallel to the foil causes a slight misorientation of the two crystal parts separated by the dislocation. The tilt angle, θ, depends on the foil thickness and on the position of the dislocation within the foil, being a maximum, $\theta_{\max} = b/t$, where t is the foil

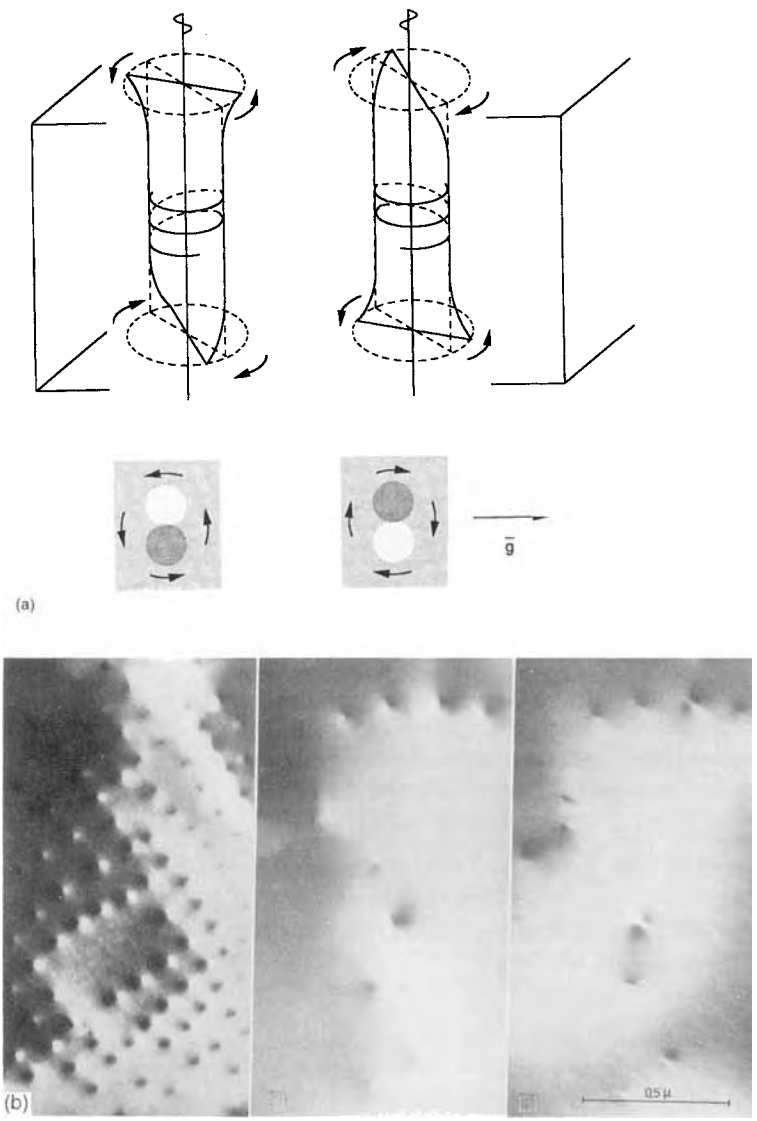


Figure 44. (a) Schematic illustration of the surface relaxation around the emergence points of screw dislocations in the foil surfaces. This surface relaxation produces image contrast for $\boldsymbol{g} \cdot \boldsymbol{b} = 0$. (b) Screw dislocations viewed end-on in platinum [59].

thickness, if the dislocation is in the central plane. As a result of this slight 'buckling' of the foil, a brightness difference is produced between the two crystal parts, separated by the dislocation. The tilt angle, θ, can be measured by the displacement of the Kikuchi lines; its sense depends on the sign of the dislocation, and therefore a knowledge of θ allows the sign of the dislocation to be determined (Fig. 45) [18–20].

An edge dislocation viewed end-on along the beam direction produces contrast because in the vicinity of the dislocation the interplanar spacing is slightly modified, $\boldsymbol{g}$ changes in length and in orientation and, consequently, the diffraction conditions also change. Along a column

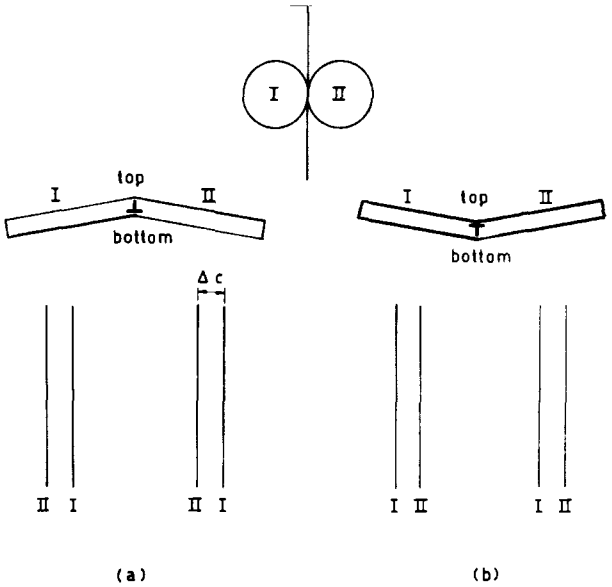


Figure 45. Buckling of a thin foil due to the presence of an edge dislocation parallel to the foil surfaces. The positions of the Kikuchi lines in areas I and II are shown schematically.

parallel to the dislocation (i.e., along z), s remains constant, but s becomes a function of x and y, chosen in the foil plane. As a result, the scattered and transmitted intensities depend on the column positions, that is an image is produced. The contours of equal s (i.e., of equal brightness) are shown in Fig. 46; they image the strain field around the edge dislocation [37, 38].

According to the simple $\boldsymbol{g} \cdot \boldsymbol{b} = 0$ criterion, an edge dislocation with its supplementary half-plane parallel to the foil plane or, stated otherwise, with its Burgers vector, $\boldsymbol{b}$, along the incident beam, would not produce any contrast. Due to the presence of the 'bump' in the glide plane (i.e., the component of the displacement field) towards the supplementary half-plane, perpendicular to the glide plane, $\boldsymbol{g} \cdot \boldsymbol{R}$ is not zero for all $\boldsymbol{g}$ vectors perpendicular to $\boldsymbol{b}$ and some contrast is produced. Prismatic dislocation loops in planes parallel to the foil plane have this configuration. The displacement field of such loops

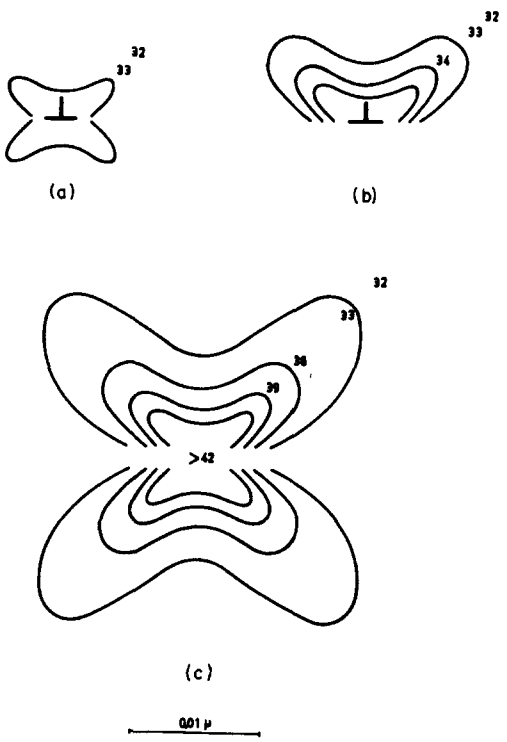


Figure 46. Contours of equal deviation parameter s in the vicinity of an edge dislocation viewed end-on [37, 38].

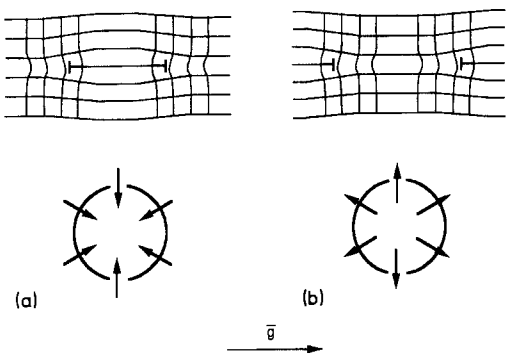


Figure 47. Radial displacement field around prismatic loops.

now contains a radial component $\boldsymbol{R}_r$ which is inward or outward for vacancy and interstitial loops (Fig. 47), respectively, as well as a normal component $\boldsymbol{R}_n = \boldsymbol{b}$.

For a diffraction vector $\boldsymbol{g}$ parallel to the foil plane the dot product with the normal component $\boldsymbol{g} \cdot \boldsymbol{R}_n = \boldsymbol{g} \cdot \boldsymbol{b}$ will be zero everywhere along the loop. However, $\boldsymbol{g} \cdot \boldsymbol{R}_r$ varies along the loop and vanishes only along the two diametrically opposite segments where $\boldsymbol{g}$ is perpendicular to $\boldsymbol{R}_r$, as represented in Fig. 47. As a result, there will be two short segments only along which complete extinction occurs; the 'line of no contrast' joining these two segments is perpendicular to the active $\boldsymbol{g}$ vector.

Somewhat against intuition, one finds that parallel dislocation lines with the same Burgers vector do not necessarily exhibit the same contrast, especially when they are close one to the other, as in a ribbon. One of the lines is usually imaged as a darker line than the other(s); which line will exhibit the strongest line contrast depends on the sign of s and on the sense of $\boldsymbol{g}$. The effect is particularly striking in triple ribbons in face-centered-cubic, low stacking fault energy alloys, and in graphite. An analytical theory, based on the kinematical diffraction theory, allows one to account satisfactorily for the observations, on noting that the total strain field of a triple ribbon is different from that resulting from the mere superposition of the strain fields of three isolated dislocations [39].

Semi-Quantitative Considerations

It is often useful to be able to predict semiquantitatively, or even only qualitatively, the two-beam image characteristics to be expected for a given defect. For dislocation lines this is possible within the framework of the dynamical theory, including anomalous absorption, by referring to the analytically soluble case of the stacking fault.

Consider, for example, an inclined edge dislocation with its Burgers vector parallel to the foil plane and an active reflection such that $\boldsymbol{g} \cdot \boldsymbol{b} = 1$. A sketch of the displacement field of the dislocation is represented schematically in Fig. 48. The bright field and dark field image profiles can be obtained by considering a row of columns along a line perpendicular to the dislocation line, and computing for each of these columns the amplitude of the transmitted and scattered beams.

We first note that these amplitudes are, to a large extent, determined by the phase relation between the top and bottom ends of the columns and not so much by the details of the variation of this phase along the column. In any case, the phase varies rapidly only in the close vicinity of the dislocation core. We therefore accept as a reasonable approximation that the amplitudes emerging from such a column will be

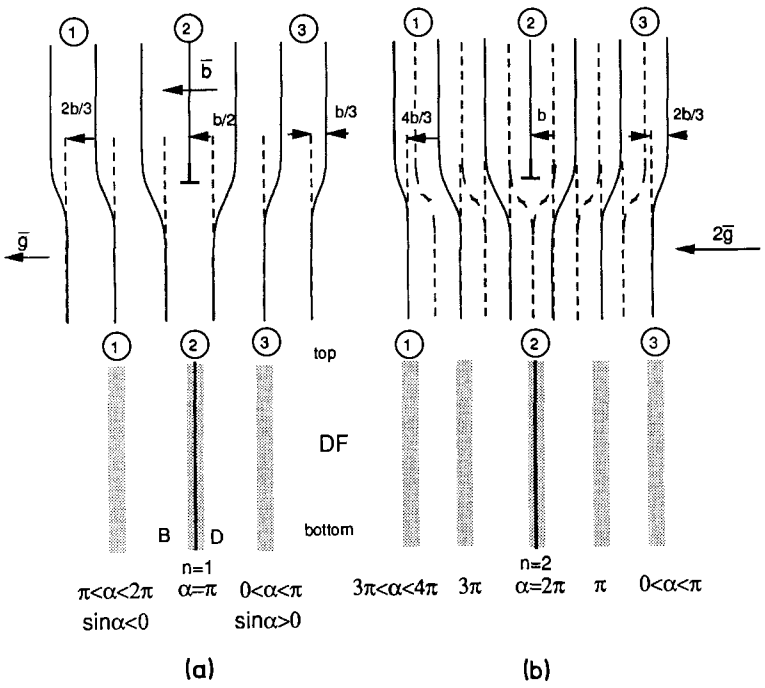


Figure 48. Schematic illustration of the displacement field around an edge dislocation for the cases $n = 1$ (a) and $n = 2$ (b), illustrating how the qualitative features of the image can be deduced from the profiles of stacking fault fringes. DF, dark field.

the same as those emerging from a column which intersects a stacking fault at the level of the dislocation core and which introduces abruptly the same shift between the top and the bottom as the dislocation. Consider as a simple example the columns passing through the dislocation core for the case $n \equiv \boldsymbol{g} \cdot \boldsymbol{b} = 1$. The phase shift between the top and the bottom of these columns is then π, as is immediately evident from the geometry shown in Fig. 48a. The inclined dislocation will then exhibit along its core the same contrast variation with depth as an inclined stacking fault with $\alpha = \pi$, situated everywhere at the same level as the dislocation core. If for the same dislocation $n \equiv \boldsymbol{g} \cdot \boldsymbol{b} = 2$, the brightness along the core will be the same as that of a perfect crystal with the same thickness, since now $\alpha = 2\pi$ along the central strip (Fig. 48b).

The image of an inclined dislocation is a two-dimensional brightness map and requires a knowledge of a large number of section profiles or, alternatively, of a number of longitudinal profiles, parallel to the dislocation. Profiles of the latter type can be obtained by considering strips of stacking fault, all parallel to the dislocation core and at the same level, but at increasing distances from this core. For $n = 1$ the central strip corresponds, as mentioned above, to $\alpha = \pi$. The corresponding α values for successive strips on moving to the right away from the dislocation vary from $\alpha = \pi$ to 2π far to the right. As a result, on the right of the dislocation, $\sin \alpha < 0$. On the left of the dislocation the α values vary from 0 at the extreme left to $\alpha = \pi$ at the dislocation position; on the left of the dislocation $\sin \alpha > 0$. The image profiles of stacking faults (for $s = 0$) show

that the contrast will oscillate. Along the strips where $\sin\alpha > 0$ (i.e., on the left of the dislocation) the first extremum behind the entrance face will be a maximum in the bright field image, whereas on the left of the dislocation, where $\sin\alpha < 0$, the first extremum will be a maximum. Near the exit face the last fringe in the bright field image will be the same as the first, on the left as well as on the right, of the dislocation. The dislocation contrast thus oscillates as a function of depth, since the maxima and minima in brightness on the two sides of the dislocation (i.e., for $\sin\alpha > 0$ and $\sin\alpha < 0$) are in antiphase. The oscillations will be most pronounced for foil thicknesses equal to $(2k+\frac{1}{2})t_g$ or $(2k+\frac{3}{2})t_g$ as for these thicknesses the stacking fault fringes in the bright field image vary strongly in brightness (see Fig. 30). In the dark field image the brightness variation is less pronounced for the same thickness; on the other hand, the variation is now more pronounced for foil thicknesses of $2kt_g$ and $(2k+1)t_g$. At the same time, the difference between the brightness for $\sin\alpha > 0$ and $\sin\alpha < 0$ is more pronounced in the bright field for thicknesses of $(2k+\frac{1}{2})t_g$ and $(2k+\frac{3}{2})t_g$ than for thicknesses of $2kt_g$ or $(2k+1)t_g$. In the former case the contrast is 'oscillating', whereas in the latter case it is 'dotted'; the inverse is true for the dark field image. The stacking fault fringe profiles suggest in the same manner that the bright field image will be similar to the dark field image close to the front surface, but quasi-complementary close to the exit surface.

As a second example, we consider the image associated with the displacement field around a spherical inclusion with $\varepsilon > 0$, as described by Eq. (46) [40–42]. Also in this case we can deduce from

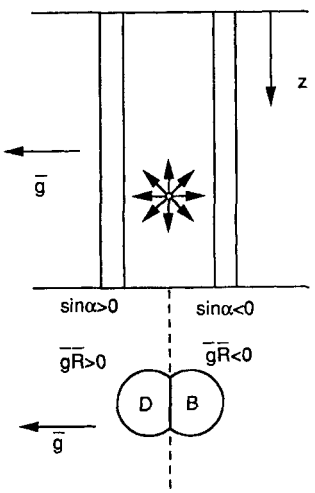


Figure 49. Strain field associated with a spherical inclusion. D, dark; B, Bright.

intuitive considerations the areas in which $\sin\alpha > 0$ and those in which $\sin\alpha < 0$, and hence conclude for a defect close to the surface (within the first extinction distance) which area will be bright and which will be dark. We consider, in particular, the spherically symmetrical displacement or strain field represented in Fig. 49. A line (or plane) of no contrast, along which $\boldsymbol{g}\cdot\boldsymbol{R} = 0$ separates two regions, one in which $\boldsymbol{g}\cdot\boldsymbol{R} > 0$ and one in which $\boldsymbol{g}\cdot\boldsymbol{R} < 0$. Since $|R|$ is small compared to a lattice vector, if $\boldsymbol{g}\cdot\boldsymbol{R} > 0$ then $\sin\alpha > 0$ and if $\boldsymbol{g}\cdot\boldsymbol{R} < 0$ then $\sin\alpha < 0$. That this is so can be deduced from the consideration that, if $\boldsymbol{g}\cdot\boldsymbol{R}$ is positive for all z values along the column, the integrated phase difference between top and bottom of the column is positive but smaller than π and hence $\sin\alpha > 0$, the fastest phase change occurring at the level of the inclusion. The brightness at the exit end of the column is then the same as that of a stacking fault, assuming the effective phase shift, α_{eff}, to occur at the level of the defect.

The value of α_{eff} decreases with increasing distance from the inclusion and changes sign along the line of no contrast. The image characteristics of stacking faults in sufficiently thick foils, close to the surface, allows us to deduce the dark field image of this kind of defect when close to the back surface. The last fringe in the dark field image of a stacking fault for which $\sin\alpha > 0$ is dark. We can thus conclude that $\boldsymbol{g}$ points towards the dark lobe for an inclusion with $\varepsilon > 0$ situated close to the back surface. Black and dark are reversed for $\varepsilon < 0$. The model also accounts for the periodic interchange with period t_g of bright and dark lobes with the depth position of the spherical inclusion.

Kinematical Theory of Dislocation Contrast

Within the framework of the kinematical diffraction theory, image profiles of dislocations are obtained by inserting the adequate expression for the displacement field $\boldsymbol{R}(\boldsymbol{r})$ in Eq. (50) and integrating along columns situated on lines normal to the dislocation line. Due to the symmetry of the displacement field of a dislocation, the profile so obtained is independent of the chosen line of columns for dislocations parallel to the surfaces of the foil.

For example, for a screw dislocation oriented along the y axis parallel to the foil plane and situated at a depth, d, behind the entrance face the displacement field is described, according to the isotropic linear elasticity theory by the expressions

$$R_x = 0 \tag{159a}$$

$$R_y = \frac{b\phi}{2\pi} \tag{159b}$$

$$R_z = 0 \tag{159c}$$

with $\phi = \operatorname{arctg}[(z-d)/x]$, that is all the displacements are parallel to $\boldsymbol{b}$. The image profile is then obtained by performing the integration

$$A(\boldsymbol{g}) = F_g \int_0^{z_0} \exp(2\pi i s_g z) \times \exp(in)\operatorname{arctg}\left(\frac{z-d}{x}\right) \mathrm{d}z \tag{160}$$

where $n \equiv \boldsymbol{g}\cdot\boldsymbol{b}$ for various values of the parameter x, and z_0 is the foil thickness. After a number of approximations, the integrals can be obtained analytically in terms of Bessel functions.

In their discussion of image profiles of dislocations, Hirsch et al. [10, 11] and Gevers [12] made extensive use of amplitude–phase (A–P) diagrams. We shall follow the same type of reasoning, since this allows us to identify more clearly the approximations and limitations of the theory. The integration along a column is represented graphically by the vector sum of the elementary contributions due to the slices $\mathrm{d}z$ along the column. In a perfect crystal we have seen that the vectors representing the amplitudes scattered by successive slices enclose a constant angle $\mathrm{d}\theta = 2\pi s\,\mathrm{d}z$, as a result of the constant phase difference between successive slices $\mathrm{d}z$. These small vectors form a regular polygon, which in the limit for $\mathrm{d}z \to 0$ becomes an arc of a circle with radius $\frac{1}{2\pi s}$. The length of the circular arc is equal to the column length, and the amplitude scattered by the column is given by the length of the vector joining the two end-points of the circular arc.

When a stacking fault is present, a discontinuous phase change $\alpha = 2\pi\boldsymbol{g}\cdot\boldsymbol{R}$ occurs in each column at the level of the stacking fault. This is reflected in the A–P

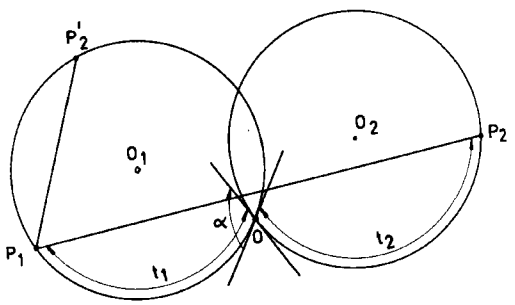


Figure 50. Amplitude–phase diagram for a column intersecting a stacking fault. The thickness of the two crystal parts are $t_1 \equiv z_1$ and $t_2 \equiv z_2$ [13, 14].

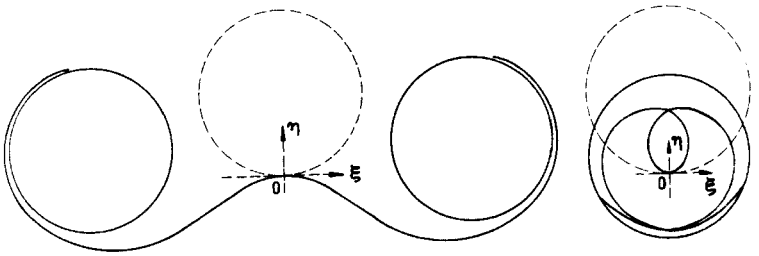
diagram by a relative rotation over an angle α of the two circles representing the A–P diagrams of the perfect parts, the tangents enclosing an angle α (Fig. 50). The amplitude diffracted by the faulted crystal is then given by the vector joining the end-points P_1 and P_2. If we choose the origin of the diagram at the position of the stacking fault, the two circular arcs have lengths equal to the front (z_1) and rear part (z_2) of the foil: $z_1 + z_2 = z_0$, where z_0 is the foil thickness. For an inclined stacking fault the end-points P_1 and P_2 corresponding to successive columns along a profile shift continuously over the same arc length in the same sense. The resulting amplitude thus varies periodically with a depth period $1/s$, describing the stacking fault fringes, and is represented analytically by Eq. (52).

The A–P diagram for a column intersecting a domain boundary also consists of two circular arcs with lengths equal to the thicknesses of front and rear parts; they join smoothly with a common tangent at the level of the boundary, but they have different radii, $\frac{1}{2\pi s_1}$ and $\frac{1}{2\pi s_2}$, since the s values are different in the two perfect parts. In addition, in this case an inclined domain boundary will produce a fringe pattern with a depth quasiperiod, which is somewhat variable over the width of the fringe pattern between $1/s_1$ and $1/s_2$. The A–P diagram is the geometrical representation of the analytical expression Eq. (53).

In the A–P diagram for a foil containing a dislocation the phase difference between successive slices of the column at levels z and $z + \mathrm{d}z$ is no longer a constant $\mathrm{d}\phi = 2\pi s\,\mathrm{d}z$, since a supplementary phase difference results from the displacements described by $\alpha(x, y, z)$. Depending on the signs of x, s, and z, this additional phase shift will either be added or subtracted; its magnitude depends on x and z and is given in the simple case of the screw dislocation by $n \cdot d[\mathrm{arctg}(z/x)]$. For $z \gg x$, this additional shift becomes zero and the final shape becomes again a circle with radius $\frac{1}{2\pi s}$ as for the perfect crystal. Close to the dislocation and for s and nx having the same sign, that is for $n\beta > 0$ (with $\beta = 2\pi s x$), the quantity $n\,\mathrm{arctg}(z/x)$ has the same sign as $2\pi s z$ and the angle between two successive vectors is now larger than $2\pi s\,\mathrm{d}z$, at least near the depth position of the dislocation, which is chosen as the origin of the diagram (i.e., of the z axis). As z becomes larger the angle again approaches $2\pi s\,\mathrm{d}z$. The resulting curve will be a wound-up spiral which gradually tends to a circle, approaching it from the interior, the circle being the limiting curve (Fig. 51b). If, on the other hand, s and nx have opposite signs (i.e., for $n\beta < 0$), $\mathrm{arctg}(z/x)$ and $2\pi s z$ have opposite signs and the resulting angle between successive vectors will be smaller than the value $2\pi s\,\mathrm{d}z$ in the perfect crystal by

$$n \cdot d[\mathrm{arctg}(z/x)] \equiv n[x/(x^2 + z^2)]\,\mathrm{d}z.$$

Again, as z becomes large, the additional phase difference tends to zero and the



(a) (b)

Figure 51. Amplitude–phase diagram for a column passing close to a dislocation core ($n = 2$) [13, 14]: (a) unwound spiral; (b) wound-up spiral.

curve approaches a circle with radius $\frac{1}{2}\pi s$. The A–P diagram is now an unwound spiral approaching the limiting circles from the outside, as shown in Fig. 51a. The scattered amplitude is again obtained by taking an arc proportional to z_1 on this curve, in the negative sense leading to P_1 and an arc proportional to z_2 in the positive sense leading to P_2. The vector $\overrightarrow{P_1P_2}$ is then proportional to the scattered amplitude for the given column, that is for a given x value. Since x has different signs on the two sides of the dislocation, the A–P diagram will be an unwound spiral on one side and a wound-up spiral on the other. The vector representing the diffracted amplitude will clearly be larger for those columns for which the distance between the centers of the two limiting circles will be the largest, that is the amplitude will be largest on that side of the dislocation where the A–P diagram is an unwound spiral. This is the side where in the bright field image a dark line will be observed, called the image side (see Sec. 1.1.18.2 of this Chapter). We note that the A–P diagram depends only on $n \equiv \boldsymbol{g} \cdot \boldsymbol{b}$ and on the product $\beta \equiv 2\pi sx$, but not on s and x separately. This is consistent with the fact that changing the sign of s changes the image side. In principle, constructing a sufficient number of A–P diagrams allows one to deduce the image profiles. It is clear that for an inclined dislocation line the length of the arcs to be taken along the spiral-shaped A–P diagrams will vary continuously with the position along the dislocation, one increasing, the other one decreasing. Hereby the end-points of the A–P diagram, which determine the scattered amplitude, will in general vary periodically as the end-points describe the limiting circles. This oscillatory behavior was suppressed in the approximation introduced by Hirsch et al. [11]. The assumption was made that the square of the separation of the centers of the limiting circles is a convenient measure of the scattered intensity. This is a reasonable assumption if s is sufficiently large so that the limiting circles acquire a small radius compared to the separation of their centers. For screw dislocations Hirsch et al. [11] obtained the computed profiles shown in Fig. 24 for different values of n. Similar calculations, using the same approximations, have been performed by Gevers [12] for perfect as well as partial dislocations of edge and mixed character. The results for pure edge dislocations are shown in Fig. 25.

The Weak-Beam Method

The limitations of kinematical theory apply to the results of the preceding

chapter and we therefore conclude that the computed image profiles are only valid for very thin foils and for large s values.

From the image profiles shown in Figs. 24 and 25, we can deduce that when s is large the same $\beta(=2\pi sx)$ value is reached for small x. This implies that for large s values the peak shift and the peak width will be small. This effect, which is consistent with the observations, is systematically exploited in the weak-beam method [43]. It allows one to obtain very well localized and sharp images of the partial dislocations in narrow ribbons, as required for the measurement of stacking fault energies. Unfortunately, with increasing s value the image contrast decreases and long exposure times are needed to record the image. In practice, a reasonable trade-off between image resolution and exposure time seems to be achieved for s values of the order of $0.2\,\mathrm{nm}^{-1}$ for 100 kV electrons.

Usually a high order reflection of $3\boldsymbol{g}$ or $4\boldsymbol{g}$ is brought in the exact Bragg position and a dark field image is made in the reflection $\boldsymbol{g}$. Alternatively, a low order reflection, $\boldsymbol{g}$ or $2\boldsymbol{g}$, may be excited and $-\boldsymbol{g}$ be used for imaging. These imaging conditions are represented in Fig. 52. In order to realize these diffraction conditions exactly, the Kikuchi pattern is of great help; moreover, it allows the measurement of s. In weak-beam images the depth period of extinction contours and of stacking fault fringes is given approximately by its kinematical value $1/s_g$. Using such large s values it is possible to image, for instance, antiphase boundaries in alloys as fringe patterns, even though the extinction distance of the superlattice reflection used is larger than the foil thickness.

The kinematical theory allows us to derive approximate expressions for the peak width and peak positions of weak-beam dislocation images [43–45]. The columns close to the dislocation core can be considered as consisting of three parts (Fig. 53). The central part contains the dislocation, and the two other parts are perfect. In the central part the lattice planes of interest are inclined with respect to their orientation in the perfect parts, in such a way that somewhere close to the dislocation core the local deviation parameter is much smaller than in the perfect part. The scattered intensity will then mainly originate from this region, producing a bright peak on a darker background in the dark field, weak-beam image. The amplitude scattered by a column at x is

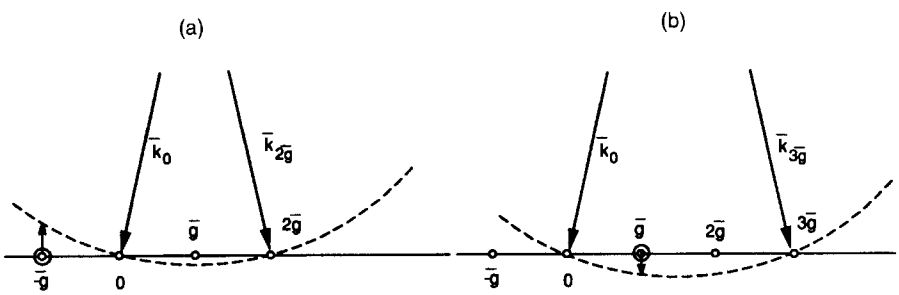


Figure 52. Imaging conditions used in the weak-beam method: (a) $s = 0$ for $2\boldsymbol{g}$, image is made in $-\boldsymbol{g}$; (b) $s = 0$ for $3\boldsymbol{g}$, image is made in $\boldsymbol{g}$ [44].

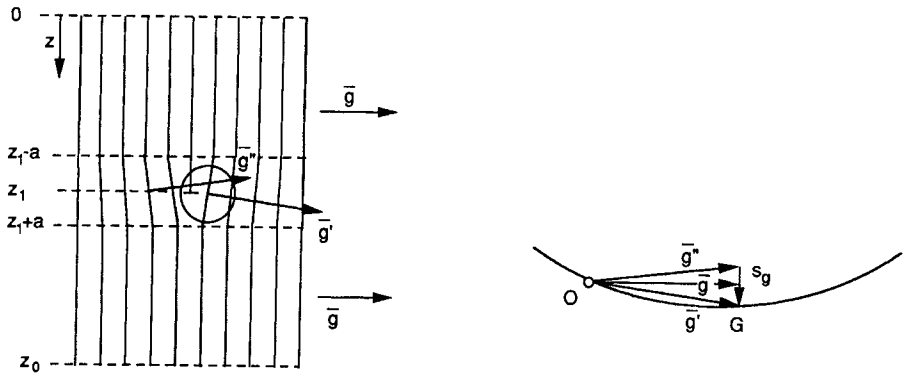


Figure 53. Foil containing an edge dislocation. Model used in discussing dislocation contrast according to the weak-beam method. The foil is assumed to consist of three lamellae: 1 $[0 \leqslant z \leqslant (z_1 - a)]$ and 3 $[(z + a) \leqslant z \leqslant 70]$ are perfect; part 2 $[(z_1 - a) < z < (z_1 + a)]$ contains the dislocation [45].

given by

$$A \propto \int_0^{z_0} \exp\left[2\pi i[s_g z + \boldsymbol{g} \cdot \boldsymbol{R}(x,z)]\right] \mathrm{d}z \quad (161)$$

This integral can be split in three parts corresponding to the three lamellae in the model shown in Fig. 53:

$$A \propto \int_0^{z_1 - a} \exp(2\pi i s_g z)\,\mathrm{d}z + \int_{z_1 - a}^{z_1 + a} \exp[2\pi i(s_g z + \boldsymbol{g} \cdot \boldsymbol{R})]\,\mathrm{d}z + \int_{z_1 + a}^{z_0} \exp(2\pi i s_g z)\,\mathrm{d}z \quad (162)$$

The first and third integral refer to the perfect parts; they do not depend on the presence of the defect. Since s is large in these parts their contribution is small. Their A–P diagrams consist of small circles with a radius $\frac{1}{2}\pi s$. These two circles are connected by a circular arc with a much larger radius, $1/s_{\mathrm{eff}}$, which is the A–P diagram of the central part. The amplitude scattered by the column is then given, to a good approximation, by the length of the segment joining the centers of the two small circles. This length is well approximated by the second integral, which we now consider (Fig. 54).

We can write the displacement function $\boldsymbol{R}(x,z)$ as a Taylor expansion in the vicinity of the core position $z = z_1$:

$$\boldsymbol{R} = \boldsymbol{R}(z_1) + (z - z_1)\left(\frac{\partial \boldsymbol{R}}{\partial z}\right)_{z_1} + \tfrac{1}{2}(z - z_1)^2\left(\frac{\partial^2 \boldsymbol{R}}{\partial z^2}\right)_{z_1} + \cdots \quad (163)$$

Retaining only the first two terms, the second integral can be written as

$$\exp\left\{2\pi i\left[\boldsymbol{R}(z_1) - z_1\left(\frac{\partial \boldsymbol{R}}{\partial z}\right)_{z_1}\right] \cdot \boldsymbol{g}\right\} \times \int_{z-a}^{z+a} \exp\left\{2\pi i\left[s_g + \boldsymbol{g} \cdot \frac{\partial \boldsymbol{R}}{\partial z}\right] z\right\} \mathrm{d}z \quad (164)$$

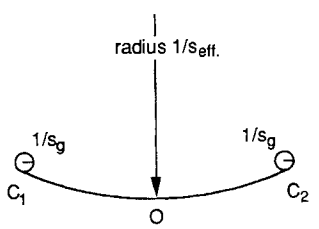


Figure 54. Amplitude–phase diagram for a column close to the dislocation core according to the weak-beam method [43].

This expression will be a maximum if the modulus of the integrandum is unity, that is for the value of x given by

$$s_g + \frac{\partial \boldsymbol{R}}{\partial z} \cdot \boldsymbol{g} = 0 \tag{165}$$

This condition is equivalent to the statement $s_{\text{eff}} = 0$ [see Eq. (61)].

Introducing the displacement field for edge and screw dislocations, adopting the FS/RH convention (see Sec. 1.1.18.2 of this Chapter) leads to the peak position $x = x_{\text{m}}$ with

$$x_m = \left(1 + \frac{K}{2(1-v)}\right)\left(-\frac{\boldsymbol{g}\cdot\boldsymbol{b}}{2\pi s_g}\right) \tag{166}$$

The parameter is $K = 1$ for an edge dislocation and $K = 0$ for a screw dislocation; ν is Poisson's ratio. In this approximation the peak position does not depend on the foil thickness or on the depth position of the dislocation. The image side, that is the sign of x_{m}, is clearly determined by the sign of the product $(\boldsymbol{g}\cdot\boldsymbol{b})s_g$.

Using the same model, the peak width at half maximum can be deduced from the kinematical approximation. For $\boldsymbol{g}\cdot\boldsymbol{b} = 2$ one finds

$$\Delta x = \frac{0.28}{|s_g|[1 + K/2(1-\nu)]} \tag{167}$$

With $\nu = \frac{1}{3}$ and $|s_g| = 0.2\,\text{nm}^{-1}$, one finds that $\Delta x \simeq 2.5\,\text{nm}$ for an edge dislocation.

With increasing value of $\boldsymbol{g}\cdot\boldsymbol{b}$ the image peak moves away from the core position. The larger the value of $\boldsymbol{g}\cdot\boldsymbol{b}$ the larger the values of s_g needed in order to achieve the same precision in the image position. In practice, this limits the values of $\boldsymbol{g}\cdot\boldsymbol{b}$ to $\leqslant 2$.

An example of a weak-beam image in the layered crystal $RuSe_2$ is reproduced in Fig. 55.

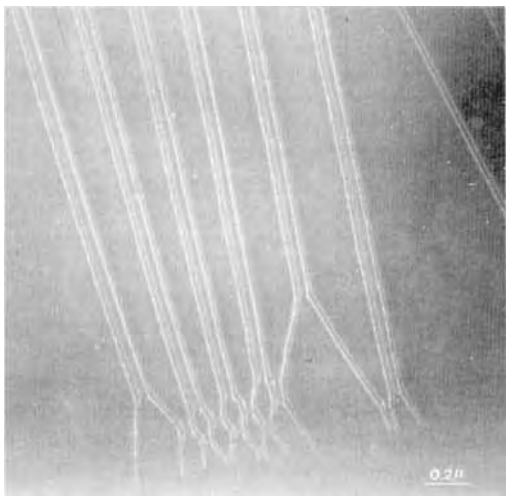


Figure 55. Weak-beam image of four-fold dislocation ribbons in $RuSe_2$. (Courtesy of J. Van Landuyt.)

The dynamical theory, neglecting anomalous absorption for simplicity, leads to essentially the same qualitative results. In terms of this theory the scattered beam, in the first part along a column close to the dislocation on the image side, oscillates with a small amplitude and with a depth period given to a good approximation by the kinematical value $1/s$ (Fig. 56). In the second part of this column, where $s_{\text{eff}} \simeq 0$, the Pendellösung oscillations acquire a large amplitude and a depth period approximated by t_g. As this second part is thin, only a fraction of an oscillation can develop, and in part three the amplitude of the oscillation and its depth period again become the same as in part one. However, the average intensity level has now become larger in part three, in particular for the column along which s_{eff} becomes zero at the level of the dislocation core. Hence the observed intensity at the exit face of that column will be larger than that for columns which are farther away from the dislocation, and will thus show up as a bright line.

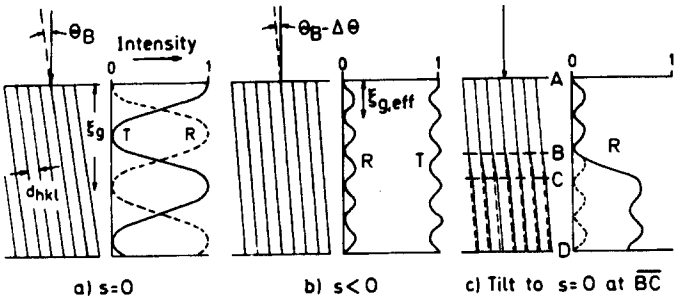


Figure 56. Weak-beam image formation at a dislocation according to the dynamical theory [43]. $\xi_g \equiv t_g$; T, transmitted beam amplitude; R, scattered beam amplitude. (a) $s = 0$; (b) $s < 0$; (c) tilt to $s = 0$ at BC.

1.1.1.18 Dislocation Contrast: Dynamical Theory

Image Simulation

One-dimensional profiles and two-dimensional maps which describe quantitatively the experimentally observed images are only obtained by applying the dynamical theory including anomalous absorption. Equations (116) or Eqs. (114) have to be integrated, with α being, in general, a function of x and z. For a screw dislocation located a distance d behind the entrance face this function becomes, for example,

$$\alpha = n\,\mathrm{arctg}[(z-d)/x] \qquad (168)$$

with $n = \boldsymbol{g}\cdot\boldsymbol{b}$; n is an integer for perfect dislocations, but it may be a fraction for partial dislocations. If d is considered to be a constant, a profile along x is sufficient to describe the image. For inclined dislocations d becomes an additional parameter and a two-dimensional map is desirable for comparison with experimental images.

Analytical solutions are difficult, if not impossible, to obtain in most cases. Numerically computed image profiles are available for a number of representative dislocation configurations and are reviewed below. A semiquantitative analytical discussion of the most striking image properties is possible [15, 45].

In principle, the computation procedure for profiles is a multislice method. It consists in considering a row of columns situated along the x axis. The integration is performed along a column (i.e., for a fixed x value) by further dividing this column in thin slices dz, each slice being considered as perfect with an s value $[s_{\mathrm{eff}} = s_g + (\boldsymbol{g}\cdot\mathrm{d}\boldsymbol{R}/\mathrm{d}z)_{x_0}]$ which depends on z. The amplitudes of the scattered and transmitted beams can be obtained by the multiplication of a succession of response matrices of the type $M(\mathrm{d}z, s_{\mathrm{eff}})$ (see Sec. 1.1.10 of this Chapter). This procedure is the implementation of the 'column approximation'.

The linear character of the Howie–Whelan system of equations [Eqs. (67) and (69)] and the fact that the displacement field of a dislocation is invariant for a translation along lines parallel to the dislocation line have been exploited by Head [47] and Humble [48] to speed up the computation procedures in order to make it possible to generate rapidly two-dimensional maps which can be compared directly with observed images (Fig. 57).

The different computer programs and subroutines required to generate two-dimensional intensity maps representing

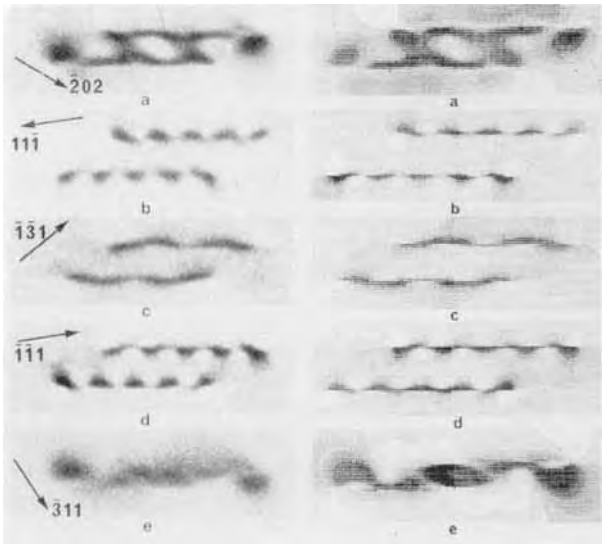


Figure 57. Examples of the quantitative agreement that can be achieved between observed and computed dislocation images: (a) observed images for different diffraction vectors; (b) corresponding computer-generated images [48].

bright and dark field images for a wide variety of single and complex defects are described in full detail by Head et al. [48]. The displacement fields of the defects are computed numerically using anisotropic linear elasticity theory. Subroutines allow one to determine the geometry of the foil, the diffraction conditions, etc., using the Kikuchi line pattern as input data. The defect identification procedure is essentially a trial-and-error method based on the inspired guess of a model based on symmetry considerations and on qualitative geometrical characteristics of the type described in previous sections in this Chapter. The model is then tested and, where necessary, further refined by quantitative comparison of the observed and computer-generated images in which a small number of parameters is varied. As the computing time is short, a comparison exercise does not require an excessive computer effort. Complete listings of the software statements are provided by Head et al. [48].

Remarkable agreement between computed and observed images can be achieved, even for complex defect configurations such as the one illustrated in Fig. 57. One of the important conclusions that has emerged from such simulations is that the extinction criterion for dislocations $\boldsymbol{g} \cdot \boldsymbol{b} = 0$ is only a first approximation and can lead to incorrect conclusions, especially in strongly anisotropic materials where the displacements around a dislocation are, in general, not parallel to $\boldsymbol{b}$, as implied in the analytical expressions based on isotropic linear elasticity.

Survey of Results of the Two-Beam Dynamical Theory [12, 13, 38]

Images of Screw Dislocations

The images for $n = 1$ and $s = 0$ for a screw dislocation parallel to the foil surfaces and located in the central plane of the foil, exhibit a single dark peak very close to

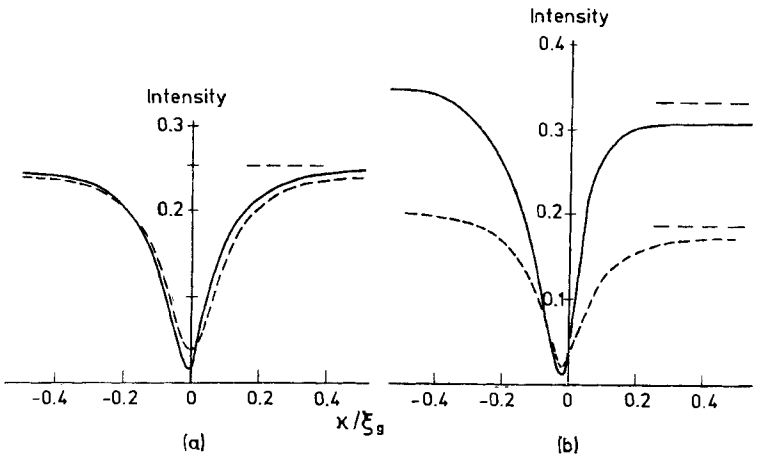


Figure 58. Image for screw dislocations in the central plane of the foil with thickness $8t_g$: (a) $n = 1$, $s = 0$; (b) $s = 0.3$. (——) Bright field image; (– – – –), dark field image [13, 14].

the position of the dislocation core, in both the bright field image and the dark field image (Fig. 58a). This is clearly in contradiction with the results of kinematical theory which predict complementary images; it is a result of anomalous absorption in thick crystals ($5t_g$ to $10t_g$). The peak width is of the order of $0.3t_g$ to $0.4t_g$.

For screw dislocations close to the surface (Fig. 59) the image becomes clearly one-sided, the image side changing periodically with depth in the crystal.

For $s \neq 0$ and sufficiently large, the dark line is displaced away from the core position in the sense predicted by the intuitive reasoning given in Sec. 1.1.1.17 of this Chapter for the bright field image. The sense of the image shift does not depend on the depth position of the dislocation, but changes with the sign of s. As a result, the image will shift continuously sideways on crossing an inclination extinction contour in the manner represented in Fig. 60.

For $s = 0$ as well as for small values of s, the image shift does depend on the depth positions z_0. Close to the surfaces the sense of the image shift alternates with a depth period t_g, but in the central part of a thick foil the image shift is small. Inclined dislocations in thick foils will thus exhibit oscillating or alternating contrast close to the foil surfaces, but not in the central part (Fig. 61). The oscillations will be in 'phase'

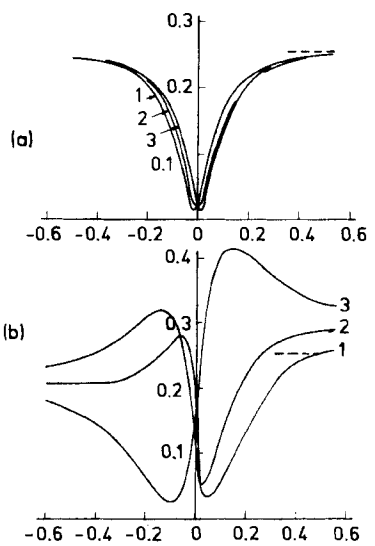


Figure 59. Bright field images for screw dislocations in foil with thickness $8t_g$ at a distance d from entrance face ($n = 1$; $z_0 = 8t_g$; $s = 0$) [13, 14]. (a) 1, $d = 4t_g$; 2, $d = 4.25t_g$; 3, $d = 4.50t_g$. (b) 1, $d = 7.25t_g$; 2, $d = 7.50t_g$; 3, $d = 7.75t_g$.

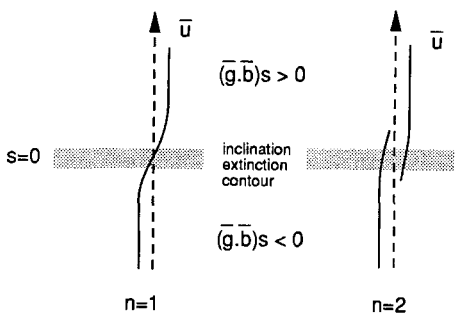


Figure 60. Image behavior on crossing an inclination extinction contour [13, 14]: (a) $n = 1$; (b) $n = 2$.

close to the entrance face in bright and dark field, but in 'antiphase' close to the exit face. The effect is a consequence of anomalous absorption and it applies to all defect images. In thick foils the bright and dark field images of defects are similar when situated close to the entrance face, but quasi-complementary close to the exit face. This has, in fact, already been pointed out for planar interfaces.

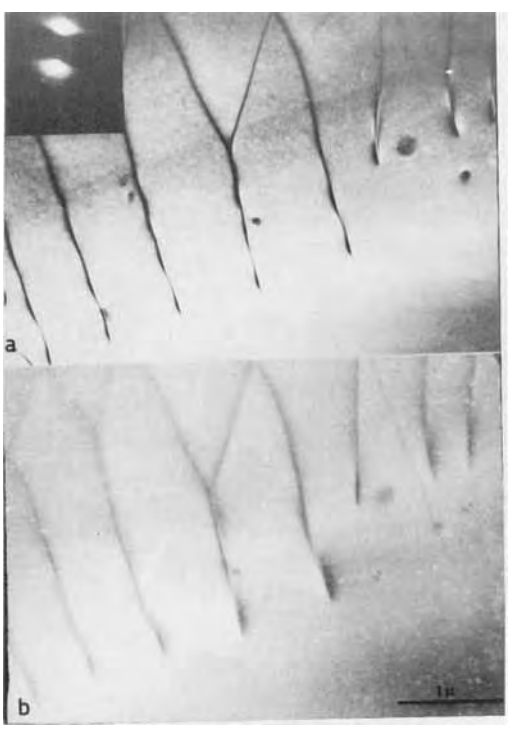


Figure 61. Dislocations in SnS_2 exhibiting oscillating contrast on approaching the surface: (a) bright field image; (b) dark field image. (Courtesy of P. Delavignette.)

This oscillating contrast can be exploited to provide a depth scale, since for $s = 0$ the oscillation period is exactly t_g. In particular, it allows a determination of the foil thickness in units of t_g, and it makes it possible to determine which end of the dislocation image corresponds to the vicinity of which foil surface (front or rear).

The occurrence of 'dotted' images at inclined dislocations can be understood intuitively, as pointed out above (see Sec. 1.1.1.17 of this Chapter) by noting that top and bottom parts of a column passing through the dislocation core are related by a phase jump of π which occurs at the level of the core. Along such columns the intensity profile for $n = 1$ will be the same as that for a stacking fault with $\alpha = \pi$. Whether predominantly 'dotted' or 'alternating' contrast occurs depends on the thickness of the foil, as does the contrast for a fault with $\alpha = \pi$. This is illustrated in Fig. 62 which shows computed bright and dark field profiles for a screw dislocation with $n = 1$ and $s = 0$ in foils with a thickness of $3t_g$ and $3.5t_g$, respectively. In the foil with a thickness of $3t_g$ the bright field image is dotted and the dark field image is alternating, whereas for a foil with a thickness of $3.5t_g$ the reverse is true.

When $n = 2$ and $s = 0$, the image exhibits two dark peaks, one on each side of the dislocation core. The two peaks are different in strength, their relative strength alternates with a period t_g with the depth in the foil. These features are illustrated in Fig. 63a; they again lead to 'oscillating'

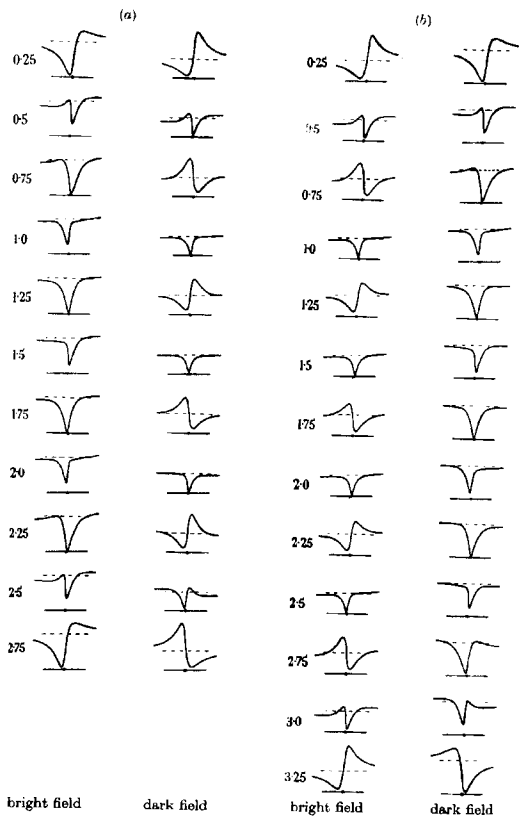


Figure 62. Image profiles of a screw dislocation ($s = 0$, $n = 1$) at different depths below the entrance face of the foil in two foils of different total thickness: (a) $3t_g$; (b) $3.5t_g$. The traces on the left-hand side are bright field images and those on the right are dark field images [13, 14].

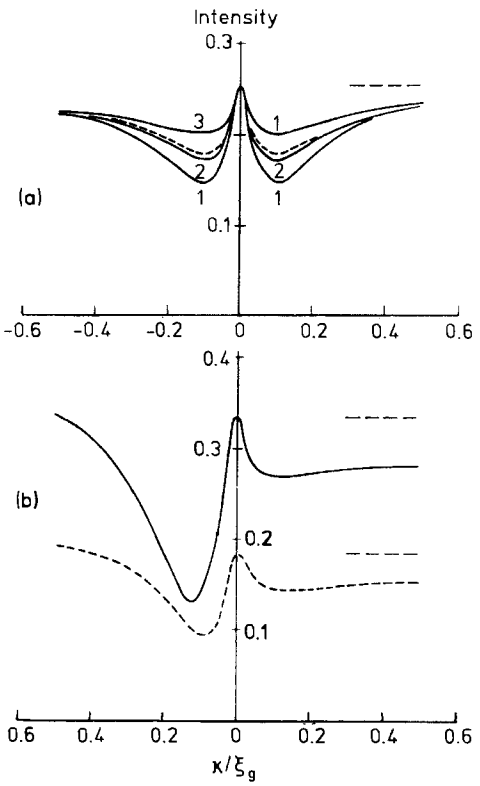


Figure 63. Image profile of a screw dislocation; foil thickness $8t_g$ [13, 14]. (a) $s = 0$, $n = 2$; curves 1–3 correspond to depths of $4t_g$, $4.25t_g$ and $4.5t_g$ behind the entrance face, respectively. (b) $s \neq 0$, $n = 2$ ($st_g = 0.3$). (——) bright field; (– – – –) dark field.

contrast at inclined dislocations for $s \simeq 0$. If $s \neq 0$ the two peaks become strongly asymmetrical, as shown in Fig. 63b, the sense of the asymmetry depending on the sign of s. Except for $s = 0$, usually only one dark peak is observed as a consequence of the asymmetry. On intersecting an equi-inclination contour with $s = 0$ the dislocation image will therefore behave as represented in Fig. 60. It is thus possible to deduce the value of n from the behavior of the image on intersecting an inclination contour. The value of n gives the projection of $\boldsymbol{b}$ on the active diffraction vector $\boldsymbol{g}$, and hence allows one to determine the length of $\boldsymbol{b}$ once its direction is known. For columns passing through the dislocation core the phase shift at the level of the core when $n = 2$ is now 2π, that is such columns will exhibit the same intensities as the perfect crystal.

Images of Edge and Mixed Dislocations

The displacement field of a mixed dislocation with a direction defined by its unit vector $\boldsymbol{u}$, parallel to the foil plane (Fig. 64) is given, according to isotropic

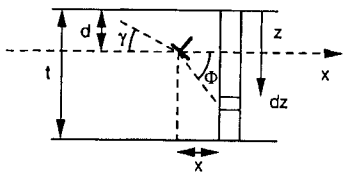


Figure 64. Reference system used to describe the displacement field due to mixed dislocations.

linear elasticity theory, by the expression

$$\boldsymbol{R} = \frac{1}{2\pi}\left\{\boldsymbol{b}\varphi + \frac{\boldsymbol{b}_e}{4(1-\nu)}\sin(2\varphi) + \left[\frac{1-2\nu}{2(1-\nu)}\ln|r| + \frac{\cos(2\varphi)}{4(1-\nu)}\right](\boldsymbol{b}\times\boldsymbol{u})\right\} \tag{169}$$

where ν is Poisson's ratio ($\nu = \frac{1}{3}$) and $\boldsymbol{b}$ is the Burgers vector, of which $\boldsymbol{b}_e$ is the edge component; $\varphi = \alpha - \gamma$; $r = (x^2 + z^2)^{1/2}$.

For a pure screw $\boldsymbol{b}\times\boldsymbol{u} = 0$ and $\boldsymbol{b}_e = 0$, and the expression reduces to $\boldsymbol{R} = (\boldsymbol{b}/2\pi)\varphi$. The term in $\boldsymbol{b}\times\boldsymbol{u}$ describes a displacement perpendicular to the slip plane towards the supplementary half-plane. The slip plane is determined by $\boldsymbol{b}$ and $\boldsymbol{u}$; it forms an angle γ with the foil plane. The character of the dislocation can be quantified by the parameter $p = (\boldsymbol{g}\cdot\boldsymbol{b}_e)/(\boldsymbol{g}\cdot\boldsymbol{b})$, which is 0 for a pure screw and 1 for a pure edge.

Computed images (Fig. 65) for a mixed dislocation with its slip plane parallel to the foil plane ($\gamma = 0$), for the following values of the parameters $n = 1$, $s = 0$, $t = 8t_g$, $z_0 = 4t_g$, and for a number of values for p, show that the image of a pure edge ($p = 1$) is wider than that of a pure screw ($p = 0$). The narrowest image is obtained for $p = -\frac{1}{2}$ and the widest for $p = \frac{1}{2}$, that is, for 45° mixed dislocations. The full width varies between $0.3t_g$ and $0.9t_g$.

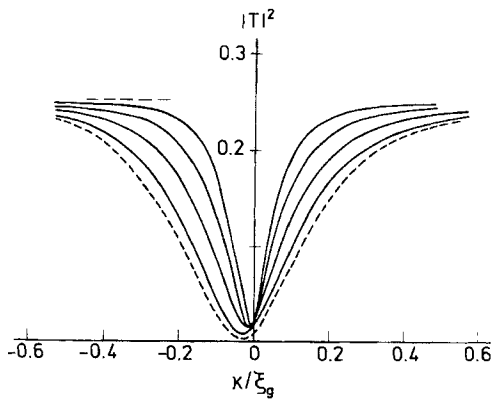


Figure 65. Computed images for mixed dislocations with slip plane parallel to the foil plane ($\gamma = 0$) and for various values of the parameter $p = (\boldsymbol{g}\cdot\boldsymbol{b}_e)/(\boldsymbol{g}\cdot\boldsymbol{b})$ [13, 14].

Even for $\boldsymbol{g}\cdot\boldsymbol{b} = 0$ a pure edge dislocation may produce contrast because of the term $\boldsymbol{b}\times\boldsymbol{u}$ in the displacement function. For a closed prismatic Frank loop parallel to the foil plane and for the imaging $\boldsymbol{g}$ vectors parallel to the loop plane $\boldsymbol{g}\cdot\boldsymbol{b} = 0$, but nevertheless complete extinction only occurs if moreover $(\boldsymbol{b}\times\boldsymbol{u})\cdot\boldsymbol{g} = 0$, which is only the case if $\boldsymbol{u}$ is parallel to the $\boldsymbol{g}$ vector. As a result, only those dislocation segments which are parallel to the acting $\boldsymbol{g}$ vector will be out of contrast. The line connecting these two segments, called the *line of no contrast*, is perpendicular to the $\boldsymbol{g}$ vector (Fig. 66). In a pure Frank loop (i.e., with its Burgers vector perpendicular to the loop plane) a line of no contrast will thus form for all $\boldsymbol{g}$ vectors parallel to the loop plane.

If the Burgers vector $\boldsymbol{b}$ is inclined with respect to the loop plane there will only be one vector $\boldsymbol{g}$ (as well as $-\boldsymbol{g}$) parallel to the loop plane, for which a line of no contrast occurs; this is the $\boldsymbol{g}$ vector perpendicular to the projection of $\boldsymbol{b}$ on the loop plane. The argument can be reversed; if among all $\boldsymbol{g}$ vectors parallel to the loop plane only one

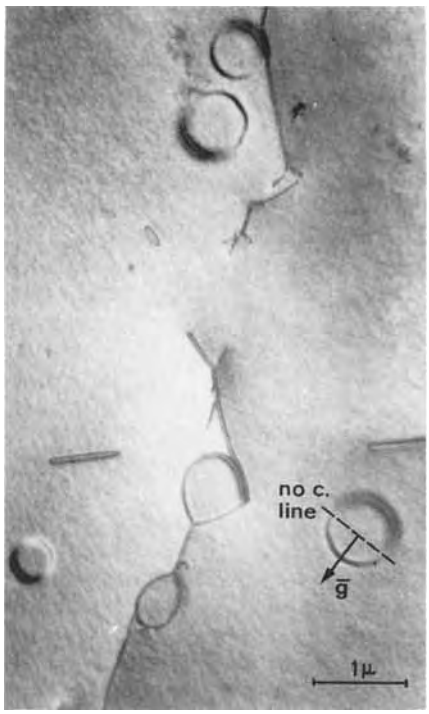


Figure 66. Images of prismatic dislocation loops with their Burgers vector parallel to the incident beam in zinc. Note the lines of no contrast perpendicular to the $\boldsymbol{g}$ vector.

produces a line of no contrast, the loop cannot be a pure Frank loop.

Changing the sign of x in Eq. (169) for the displacement field changes the sign of φ, but since all terms in Eq. (169) are even functions of either φ or x, we conclude that the image profile must be symmetrical in x. In the computed profiles shown in Fig. 67, only the half corresponding to $x > 0$ is represented; they show that for certain depth positions the image may exhibit two broad dark lines, as for instance in Fig. 66.

Images of Partial Dislocations

Since the Burgers vectors of partial dislocations are not lattice vectors, the image

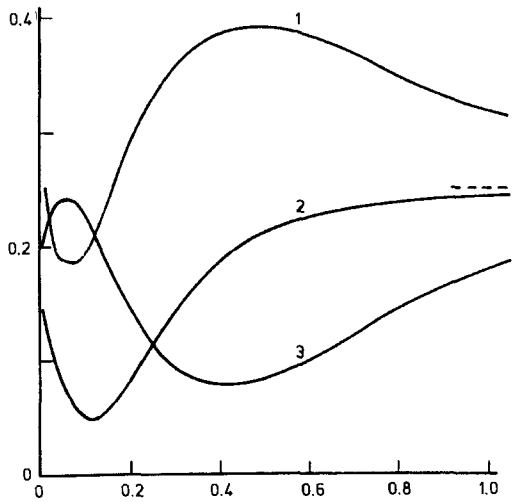


Figure 67. Computed image profiles (B.F.) for edge dislocations with their Burgers vector parallel to the incident beam. Only half of the profile is shown; it is symmetrical in x [13, 14].

order $n \equiv \boldsymbol{g} \cdot \boldsymbol{b}$ may become fractional. For instance, for Shockley partials in face-centered-cubic crystals the Burgers vector is $\frac{1}{6}[11\bar{2}]$ and the value of n becomes a multiple of $\frac{1}{3}$.

Partial dislocations form the border of stacking faults. The image profile is therefore complicated by the fact that it separates two areas, one of which has the brightness of a perfect region and the other has the contrast of a faulted area at the depth level of the partial dislocation; these brightnesses are, in general, different.

Image profiles have been computed for $n = \pm\frac{1}{3}, \pm\frac{2}{3}$, and $\pm\frac{4}{3}$. For $n = \pm\frac{1}{3}$ no visible line image is formed since the profile constitutes a continuous transition between the two brightness levels. However, for $n = \pm\frac{2}{3}$ the image consists of a dark line with a small visibility. Images with $n = \pm\frac{4}{3}$ are expected to consist of a dark line comparable to that of an ordinary dislocation (Fig. 68). Partial dislocations of the Frank type have pure edge

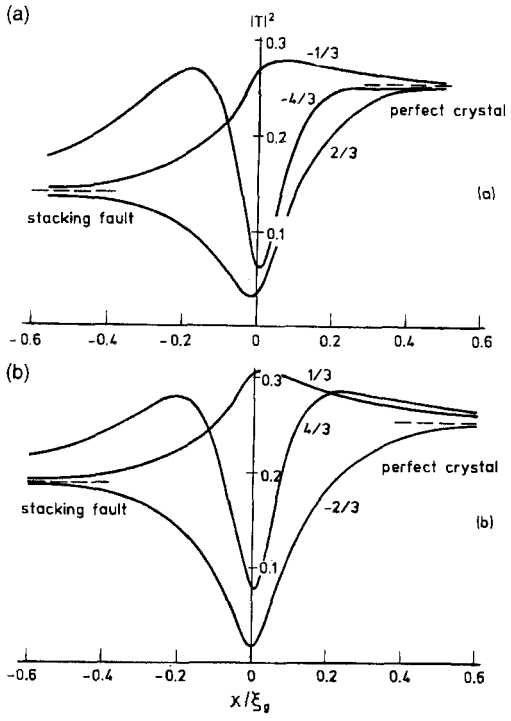


Figure 68. Image profiles (B.F.) for partial dislocations. The n values are indicated [13, 14].

character; their behavior was discussed in Sec. 1.1.1.18 of this Chapter.

Images of Dislocation Ribbons

The images of ribbons of partial dislocations [49] deserve some special consideration. Since the ribbon width is a measure of the energy of the enclosed stacking fault, the exact separation of the partial dislocations needs to be known accurately in order to make possible precise measurements of the stacking fault energy.

The image of a ribbon is *not* the superposition of the images of the separate partial dislocations; it is caused by the strain field of the ribbon, which is obtained in the framework of linear elasticity as the superposition of the strain fields of the two

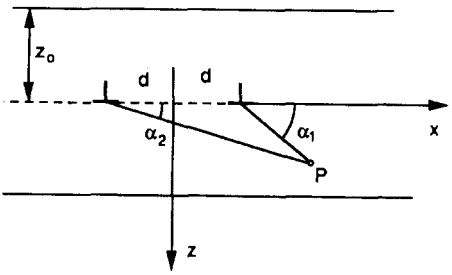


Figure 69. Reference system used in describing the displacement field of a dislocation ribbon [39].

partial dislocations. For a screw ribbon consisting of two Shockley partial dislocations, enclosing an angle of 60°, and for an active diffraction vector oriented along the bisector of the acute angle between Burgers vectors $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$, the ribbon behaves to a good approximation, as far as the contrast is concerned, as if it consisted of two screws with n values ($n = \boldsymbol{b} \cdot \boldsymbol{g}$) which are either both $+1$ or both -1. This is due to the fact that the edge components are perpendicular to $\boldsymbol{g}$ and, therefore, produce residual contrast only.

For the geometry shown in Fig. 69 the phase shift α caused by the ribbon can be formulated as

$$\alpha = n_1\alpha_1 + n_2\alpha_2 \tag{170}$$

with

$$\alpha_1 = \operatorname{arctg}\left(\frac{z - z_0}{x - d}\right) \tag{171a}$$

$$\alpha_2 = \operatorname{arctg}\left(\frac{z - z_0}{x + d}\right) \tag{171b}$$

and

$$n_1 = \boldsymbol{g} \cdot \boldsymbol{b}_1 \quad \text{and} \quad n_2 = \boldsymbol{g} \cdot \boldsymbol{b}_2 \tag{171c}$$

Such a ribbon will produce a symmetrical image when $n_1 = \pm 1$ and $n_2 = \mp 1$, since changing x into $-x$ leads to changing α_1

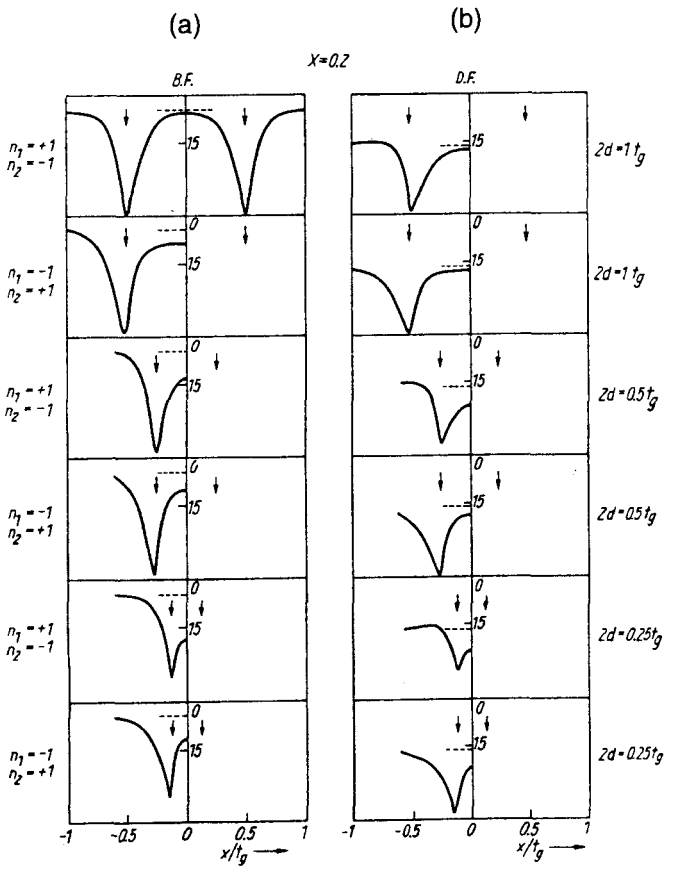


Figure 70. (a) Bright field and (b) dark field image profiles for ribbons of varying widths ($s_g t_g = 0.2$) and varying combinations of $n_1 = \pm 1$, $n_2 = \mp 1$. The profiles are symmetrical; only one half is shown. BF and DF profiles are represented. The positions of the partial dislocations are indicated by arrows [39].

into $-\alpha_2$, and vice versa, and thus $\alpha(x)$ into $\alpha(-x)$. Integrating along columns at $-x$ and at $+x$ leads to the same result, and the image is thus symmetrical.

If we change simultaneously $n_1 = n_2 = +1$ into $n_1 = n_2 = -1$ and x into $-x$ the expression for α remains unchanged. We conclude from this that the profile for $n_1 = n_2 = -1$ is the mirror image of that for $n_1 = n_2 = +1$. It is thus sufficient to discuss one of these two cases. Bright and dark field image profiles for screw ribbons are reproduced in Fig. 70 for various sign combinations of $n_1 = \boldsymbol{g} \cdot \boldsymbol{b}_1 = \pm 1$ and $n_2 = \boldsymbol{g} \cdot \boldsymbol{b}_2 = \pm 1$. In all cases the foil had a thickness of $8t_g$; the ribbon was at a depth $3.5t_g$ behind the entrance foil and $st_g = 0.2$.

When n_1 and n_2 are opposite in sign, as in Fig. 70a, the profiles are symmetrical; only one-half is therefore shown; the complete profile is obtained by a mirror operation. The positions of the dislocations are indicated by arrows. The ribbon widths are clearly different from the peak separations, due to the fact that the image displacements are in the opposite sense for the two dislocations as a result of the sign difference of n. For the sign combination $n_1 = +1$ and $n_2 = -1$ the apparent (observed) width of the ribbon is smaller than the real width; the opposite is true for

$n_1 = -1$ and $n_2 = +1$. This width difference is also found to increase with increasing s_g, since the image displacements increase with s_g. The effect of a decreasing separation, $2d$, on the contrast in the central part of the ribbon is clearly visible; the background intensity being represented by a horizontal dotted line. The center of the image corresponds to a column for which the phase shift is given by

$$\alpha_0 = (n_1 + n_2)\phi - k\pi \qquad (172)$$

where $\phi = \text{arctg}[(z - z_0)/d]$. The brightness in the center is thus the same as for a stacking fault at $z = z_0$, with α_0 as a phase shift.

If, on the other hand, n_1 and n_2 have the same sign the image sides are the same for the two partial dislocations and the peak separation in the image is more representative of the real width of the ribbon (Fig. 71b). However, the two partial dislocations are now imaged as lines of different width and brightness for $s \neq 0$. This can be understood by noting that the displacements associated with the two partial dislocations are additive outside of the ribbon, but subtractive in the region between the two partials (i.e., inside the ribbon). The strongest line image is formed outside the ribbon on the image side of the first partial dislocation. For the second partial dislocation the image side is the same as for the first one, but this is now inside the ribbon where the displacements are subtractive and hence the peak is smaller. Changing the sign of s changes the image side for both partial dislocations. The strongest image will again be outside the ribbon but on the other side, since this is now where the displacements are again additive, albeit in the opposite sense. The brightness in the central column is now the same as the background, since $\alpha(x = 0) = k\pi$. Some of these features can be observed in the computed profiles shown in Fig. 71b.

The symmetrical triple ribbons in graphite are formed by three partial dislocations with the same Burgers vector,

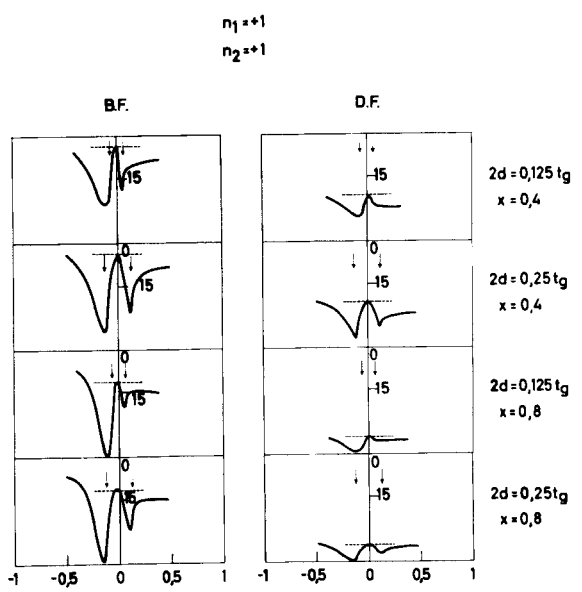


Figure 71. (a) Bright and (b) dark field image profiles for dislocation ribbons. Foil thickness t_g; $x \equiv st_g$; $z_0 = 3.5t_g$; $n_1 = n_2 = \pm 1$ [39].

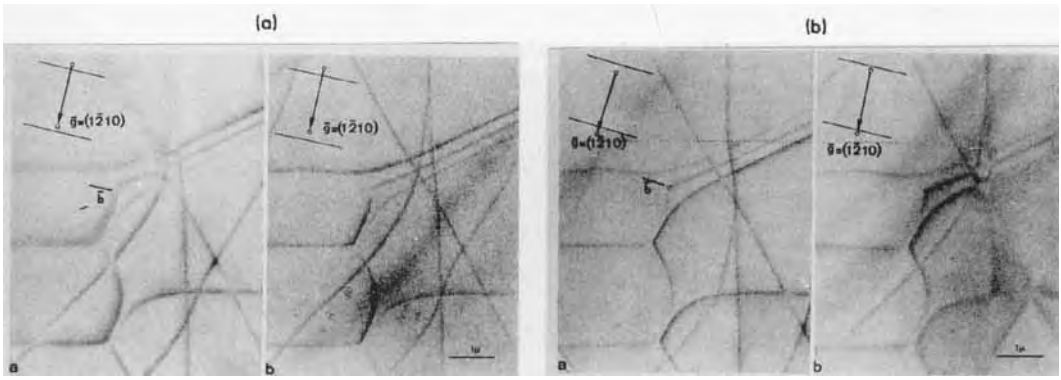


Figure 72. Image of triple ribbons in graphite under different diffraction conditions (i.e. for different s values): (a) $s > 0$; (b) $s < 0$ [39].

separating two fault ribbons (Fig. 72). As a result, the n values are always the same and hence the image sides for the three partial dislocations, which are determined by the sign of ns, are also the same. If the displacements are, for instance, additive outside the three-fold ribbon and to the left of it, they are partly additive inside the left fault ribbon, partly subtractive inside the right fault ribbon, and completely subtractive (i.e., additive in the opposite sense) outside the triple ribbon and on the right of it. As a result the peaks marking the three partial dislocations will decrease in magnitude from left to right if the image side is on the left. Changing the sign of s or changing the image side for the three partial dislocations will invert the sense in which the magnitudes of the peaks decrease. This feature can be observed in the sequence of images shown in Fig. 72.

Dislocation Dipoles

Dislocation dipoles [48] consist of two parallel dislocations with opposite Burgers vector. If the dislocations are restrained to remain in their respective glide planes they take up a stable configuration which minimizes the elastic energy. For two edge dislocations the regions of expansion in one dislocation tend to overlap the regions of compression in the other one. The configuration is such that the plane formed by the two parallel dislocations encloses an angle ϕ with their slip planes. In the case of two pure edges $\phi = 45°$.

As in the case of dislocation ribbons, the image of a dipole is not the superposition of the images of the two separate dislocations. The superposition must be carried out at the level of the strain fields in the framework of the linear elasticity theory. It has been shown by Head et al. [48] that the bright field image of an inclined dipole has a center of symmetry. This symmetry property allows one to distinguish between an image of a narrow dipole and an image of a single dislocation.

The Image Side of Dislocations

From the dynamical image simulations discussed in Sec. 1.1.1.18 of this Chapter, we concluded that the black line image of a dislocation in the bright field image is systematically one-sided provided that s is large enough. This is true for $\boldsymbol{g} \cdot \boldsymbol{b} = 1$

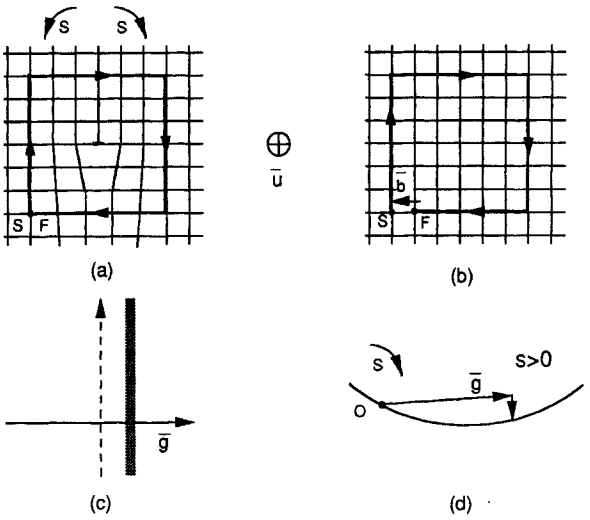


Figure 73. Illustration of the FS/RH convention for defining the sense of the Burgers vector of a dislocation: (a) real crystal; (b) perfect reference crystal; (c) relative position of image and dislocation; (d) diffraction conditions.

as well as for $\boldsymbol{g} \cdot \boldsymbol{b} = 2$. This behavior is different for $n = 1$ and $n = 2$ if $s \simeq 0$, as discussed in Sec. 1.1.1.18 of this Chapter.

The image side (i.e., the position of the black line in the bright field image on a positive print with respect to the dislocation position) is correctly given by the intuitive kinematical considerations discussed in Sec. 1.1.1.17 of this Chapter. According to this theory the image side is on that side of the dislocation core where locally the lattice planes normal to $\boldsymbol{g}$ are rotated towards the exact Bragg orientation; in Fig. 73 this is in the sense S. Finding S requires a knowledge of the sign of s; this can be determined by means of the Kikuchi line pattern, as shown in Sec. 1.1.1.3 of this Chapter. For the edge dislocation shown in Fig. 73 the positive sense (i.e., the unit vector $\boldsymbol{u}$) was chosen as entering the plane of the drawing. The Burgers vector $\boldsymbol{b}$ is determined according to the FS/RH convention as follows. A right-handed closed Burgers circuit when looking along $\boldsymbol{u}$ is constructed in the real crystal. In the perfect reference crystal the corresponding circuit is constructed and the Burgers vector $\boldsymbol{b}$ is found as the closure failure of this circuit, joining the final point to the starting point $\boldsymbol{b} = \mathrm{FS}$. For the concrete situation shown in Fig. 73, $s > 0$ and $(\boldsymbol{g} \cdot \boldsymbol{b})s < 0$; the image is indicated by a solid line and the dislocation line by a dotted line. The rule can be formulated as follows: the image side is to the right looking along the positive sense if $(\boldsymbol{g} \cdot \boldsymbol{b})s < 0$. Changing the sign of one of the three parameters $\boldsymbol{g}$, $\boldsymbol{b}$, or s changes the image side.

It should be noted that the descriptions given in different reviews are sometimes confusing and do not always agree because some authors refer to the image as seen along the incident electron beam, whereas other formulations refer to the image as seen from below. The sense of $\boldsymbol{u}$ depends on whether the first or the second viewpoint is adopted, but this changes the sign of $p = (\boldsymbol{g} \cdot \boldsymbol{b})s$. The most direct way is to apply intuitive reasoning, correctly taking into account possible electron optical image rotations.

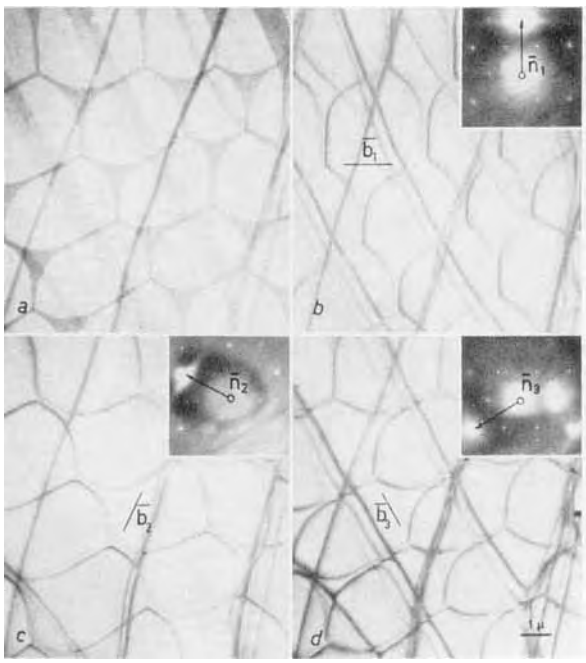


Figure 74. Dislocation network in graphite imaged under four different diffraction conditions leading to stacking fault contrast in (a) and to extinctions in (b)–(d). Note that the triple ribbon loses contrast completely in (c) [56].

Characterizing Dislocations

A full description of a dislocation line requires a determination of its core geometry and its Burgers vector, that is of the direction, magnitude and sense of $\boldsymbol{b}$ [13, 40]. Methods are available to obtain all these elements. The precise *position* of the dislocation can be found by making two images leading to opposite image sides either for active diffraction vectors $+\boldsymbol{g}$ and $-\boldsymbol{g}$ for the same sign of s, or for $+s$ and $-s$ for the same $\boldsymbol{g}$ vector. The true position of the dislocation is then between the two images.

The *direction* of the Burgers vector is determined by looking for two diffraction vectors $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ for which the dislocation is out of contrast or for which a residual contrast characteristic of $\boldsymbol{g} \cdot \boldsymbol{b} = 0$ is produced. The Burgers vector then has a direction parallel to $\boldsymbol{g}_1 \times \boldsymbol{g}_2$. An example of the application of this method to a hexagonal network of dislocations in graphite is shown in Fig. 74. In this particular case a single 'extinction' is in fact sufficient since the dislocations were known to be glide dislocations and thus have their Burgers vector in the c plane. As the foil is prepared by cleavage it is also limited by c planes. The three families of partial dislocations are seen to be successively brought to extinction using the $\boldsymbol{g}$ vectors indicated. Note also the simultaneous extinction of the three partial dislocations in the triple ribbon, showing that they have the same Burgers vector. Their contrast is nevertheless different for the different partials, as discussed in Sec. 1.1.1.18 of this Chapter.

It should be remembered that in highly anisotropic materials the simple extinction criterion $\boldsymbol{g} \cdot \boldsymbol{b} = 0$ is no longer strictly valid, as discussed in Sec. 1.1.1.18 of this

Chapter. In the case of graphite, just mentioned, the extinctions can be observed unambiguously, even though graphite is highly anisotropic. However, due to the presence of the six-fold symmetry axis along c, the c plane behaves effectively as elastically isotropic. This is also the case for dislocations in the (111) planes of face-centered-cubic crystals, due to the three-fold symmetry.

If complete extinction cannot be achieved, one should look for the weakest contrast conditions and deduce from these a plausible Burgers vector, taking the crystal structure into account. Image simulations for various $\boldsymbol{g}$ vectors, based on this Burgers vector, can then be compared with the observed images.

The *magnitude* of the Burgers vector for a perfect dislocation can be determined once its direction is known, by looking for diffraction vectors for which $\boldsymbol{g} \cdot \boldsymbol{b} = 2$. Use is made of the typical contrast effect that occurs where the dislocation crosses a bent extinction contour (see Sec. 1.1.1.18 of this Chapter). If such a diffraction vector is identified we know the length of the projection of $\boldsymbol{b}$ on $\boldsymbol{g}$. With a knowledge of the direction of $\boldsymbol{b}$ and of its projected length on $\boldsymbol{g}$, the length of $\boldsymbol{b}$ can be found.

Finally, the sense of $\boldsymbol{b}$ is found from the image side which defines the sign of $(\boldsymbol{g} \cdot \boldsymbol{b})s$. Knowing the sign of s from the Kikuchi pattern, the image side allows one to find the sign of $\boldsymbol{g} \cdot \boldsymbol{b}$. Knowledge of $\boldsymbol{g}$ then leads to the sense of $\boldsymbol{b}$.

An important application of the sign determination of the Burgers vector consists in determining whether a Frank loop is due to the precipitation of vacancy or interstitial loops, that is whether $\boldsymbol{b}$ is either $+\frac{1}{3}[111]$ or $-\frac{1}{3}[111]$. Applying the relation determining the image side to the loop represented in Fig. 75 it follows that for a loop the image is either completely inside or completely outside the dislocation ring, depending on the sign of $(\boldsymbol{g} \cdot \boldsymbol{b})s$, and since $\boldsymbol{b}$ is different for a vacancy loop and an interstitial loop, so will be the image side for the same $\boldsymbol{g}$ and s. The type of contrast experiment required for an analysis of the nature of loops is illustrated in Fig. 76. A difficulty arises because of the need to know the sense of inclination of the loop plane. If the loops are known to be parallel to the foil surface a known slope can be imposed by mounting the sample on a wedge. However, this method is not always possible.

Assuming the sense of the slope to be as represented in Fig. 75 and $\boldsymbol{g}$ and s to be as

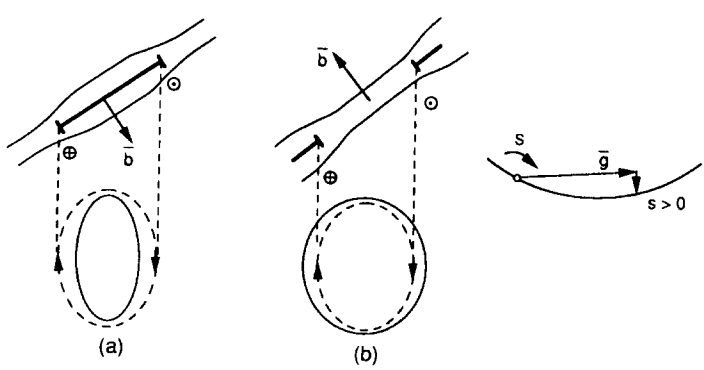


Figure 75. Image contrast of dislocation loops: (a) image inside the loop; (b) image outside the loop.

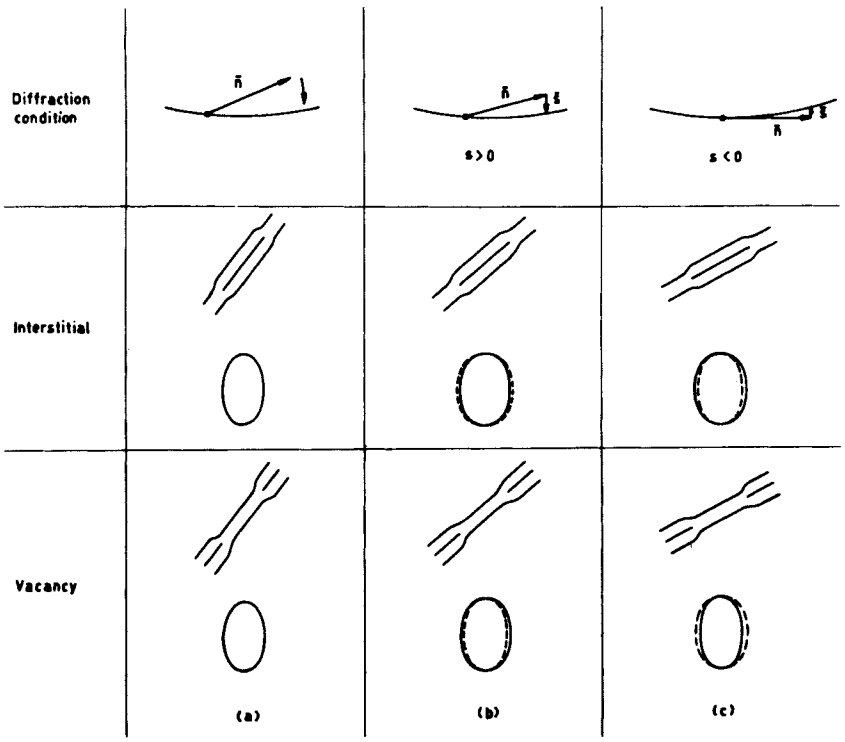


Figure 76. Contrast experiment for determining the nature of dislocation loops. Diffraction vector $\boldsymbol{g} \equiv \boldsymbol{n}$. The foil is tilted in the sense indicated.

there shown, it is evident that the image is inside for interstitial loops, whereas it is outside for vacancy loops. Changing the sign of s by tilting allows one to find the image side and hence to distinguish between the two cases.

An alternative application of the same principle consists in rotating the specimen through the exact Bragg orientation from $s > 0$ to $s < 0$ for a given $\boldsymbol{g}$ (Fig. 76). It is then found that an *interstitial* loop will grow in size because of two effects: (i) the projected size increases and (ii) the image goes from inside for $s > 0$ to outside for $s < 0$. A *vacancy* loop will grow as long as $s > 0$ because of the geometrical effect; but beyond $s = 0$ the image side changes and the image size shrinks. The experiment must clearly be performed starting with loops which are steeply inclined. One can also make use of the asymmetrical image contrast, consisting of a line of no contrast; separating a bright and a dark lobe (or crescent), characteristic of Frank loops seen end-on, which moreover are close to the surface. In the dark field image the asymmetry is the same at the top and bottom of the foil, due to anomalous absorption. If the diffraction vector $\boldsymbol{g}$ is parallel to $\boldsymbol{b}$ and points from the bright to the dark lobe in the image, the loop has interstitial character. If $\boldsymbol{g}$ points from the dark to the bright lobe, the loop is a vacancy loop. A restriction is that the loop must be close to the surface (i.e., within $\frac{1}{2} t_g$). To demonstrate that the latter

condition is satisfied, stereo images are required.

1.1.1.19 Moiré Patterns [50–52]

Intuitive Considerations

Electron microscopic specimens consisting of two similar thin films superimposed with a small orientation difference produce interference patterns consisting of parallel fringes when a $\boldsymbol{g}$ vector parallel to the film plane is excited in the two components of the sandwich. In the bright field image this fringe pattern results from the interference between the doubly transmitted and the doubly scattered beams, which enclose a small angle. This angle is usually revealed in the diffraction pattern by a doubling of the spots. A 'doubly transmitted beam' is a beam which is transmitted through the first thin film and subsequently through the second; the doubly scattered beam is formed in a similar way.

In a number of cases the geometrical features of such fringes provide useful information. A geometrical analog, consisting of the superposition of two line patterns (the lines representing lattice planes), one of them containing a dislocation, is shown in Fig. 77. In (a) the directions of the lines are the same, but their spacing is slightly different. In (b) the two patterns have the same line spacing, but the directions of the lines enclose a small angle. The moiré pattern or the superposition pattern shows a magnified representation of a dislocation [51]. Moiré patterns can thus provide 'geometrical' magnification, which was especially useful at a time when atomic resolution was not possible. With the development of atomic resolution microscopy, moiré imaging lost most of its importance; however, the geometrical features are still useful [52].

Theoretical Considerations

Consider a composite crystal consisting of two plan parallel slabs I and II [50]. Let part II of a column be derived from part I by the displacement field $\boldsymbol{u}(\boldsymbol{r})$. The phase shift between the waves diffracted locally

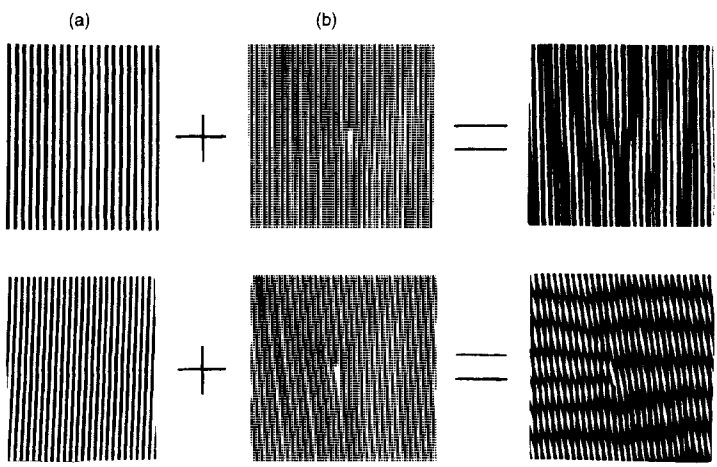


Figure 77. Geometrical analog illustrating the formation of moiré patterns: (a) parallel moiré; (b) rotation moiré. One of the two superimposed foils contains a dislocation.

by the two parts of the crystal is then $\alpha = 2\pi(\boldsymbol{g}+\boldsymbol{s})\cdot\boldsymbol{u} \simeq 2\pi(\boldsymbol{g}\cdot\boldsymbol{u})$. This expression is of the same form as the phase shift introduced by a stacking fault, the main difference being that $\boldsymbol{u}$ is *not* a constant vector $\boldsymbol{R}$ but now depends on $\boldsymbol{r}$, and hence α is also a function of $\boldsymbol{r}$. The transmitted and scattered amplitudes are then given by Eqs. (113a) and (113b), in which α enters through the periodic factor $\exp(i\alpha)$. Without solving the system of equations it is clear that the loci of the points of equal intensity (i.e., the fringes) are given by $\exp(i\alpha) = \text{Constant}$, that is by $\alpha = \text{Constant} + k2\pi$ (where k is an integer).

Assuming $\boldsymbol{r}$ to be a lattice vector $\boldsymbol{g}\cdot\boldsymbol{r} = \text{Integer}$, for small difference vectors

$$\Delta\boldsymbol{g}\cdot\boldsymbol{r} + \boldsymbol{g}\cdot\Delta\boldsymbol{r} = 0$$

$$\text{with} \quad \Delta\boldsymbol{r} = \boldsymbol{u}(\boldsymbol{r}) \tag{173}$$

and thus

$$\alpha \equiv 2\pi\boldsymbol{g}\cdot\boldsymbol{u} = -2\pi\Delta\boldsymbol{g}\cdot\boldsymbol{r} \tag{174}$$

Provided $\boldsymbol{u}(\boldsymbol{r})$ is such that $\Delta\boldsymbol{g}$ does not depend on $\boldsymbol{r}$, which is true for moiré patterns, the lines of equal intensity are given by $\Delta\boldsymbol{g}\cdot\boldsymbol{r} = \text{Constant} + k$, where k is an integer. This equation represents a set of parallel straight lines perpendicular to $\boldsymbol{K} = -\Delta\boldsymbol{g}$, where $\boldsymbol{K}$ can be considered as the wavevector of the moiré fringe system, with wavelength $\Lambda = 1/K$.

In the case of a *rotation moiré*, $K = 2g\sin(\theta/2) \simeq g\theta$ (for small θ); or, expressed in terms of the interplanar spacing d_g of the active reflection

$$\Lambda_r = d_g/\theta \tag{175}$$

The fringes are parallel to $\boldsymbol{g}$ for small θ.

For *parallel moiré patterns*, $\Delta\boldsymbol{g} = \boldsymbol{g}_2 - \boldsymbol{g}_1$ with $\boldsymbol{g}_2$ parallel to $\boldsymbol{g}_1$ or, in terms of interplanar spacings,

$$\Delta g = \frac{1}{d_2} - \frac{1}{d_1} = \frac{d_1 - d_2}{d_1 d_2} \tag{176}$$

and

$$\Lambda_p = \frac{d_1 d_2}{d_1 - d_2} \tag{177}$$

The fringes are again perpendicular to $\Delta\boldsymbol{g}$, that is they are also perpendicular to $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$.

If an orientation difference as well as a spacing difference is present, mixed moiré patterns are formed. One can always decompose $\Delta\boldsymbol{g}$ into components perpendicular and parallel to $\boldsymbol{g}$:

$$\Delta\boldsymbol{g} = \Delta\boldsymbol{g}_{\parallel} + \Delta\boldsymbol{g}_{\perp} \tag{178}$$

As the fringes are still perpendicular to $\Delta\boldsymbol{g}$, they enclose an angle β with the direction of $\boldsymbol{g}$ given by $\mathrm{tg}\,\beta = \Delta g_{\perp}/\Delta g_{\parallel}$ and

$$|\Delta g|^2 = |\Delta g_{\perp}|^2 + |\Delta g_{\parallel}|^2 \tag{179}$$

and hence

$$1/\Lambda^2 = 1/\Lambda_{\parallel}^2 + 1/\Lambda_{\perp}^2 \tag{180}$$

The intensity variation of the fringe system can be found in a similar way as for stacking faults. For a quantitative theory of the intensity profiles, see [50].

From the theory one can conclude that the positions of the moiré fringes depend on the total thickness of the sandwich, and hence the fringe positions are influenced by surface steps. Furthermore, the fringe positions depend on the specimen tilt.

We have seen that for coherent domain boundaries $\Delta\boldsymbol{g}$ is perpendicular to the interface. When such an interface is perpendicular to the incident beam, which is the usual geometry for moiré patterns, the projection of $\Delta\boldsymbol{g}$ onto the interface thus vanishes and no moiré fringes are formed.

Fringe patterns imaging this type of interface therefore have a different origin. The image for an inclined domain boundary consists of the δ fringes described above, which are perpendicular to the projection of $\Delta \boldsymbol{g}$ onto the foil plane (see Sec. 1.1.1.13 of this Chapter), that is onto the intersection lines of the interface with the foil surfaces.

If $\Delta \boldsymbol{g}$ has an arbitrary orientation with respect to the contact plane between the two crystal parts, it has a perpendicular component as well as a parallel component with respect to the interface and the image can be a complicated mixture of both types of image. The parallel component gives rise to moiré type fringes, and the perpendicular component to δ-type fringes.

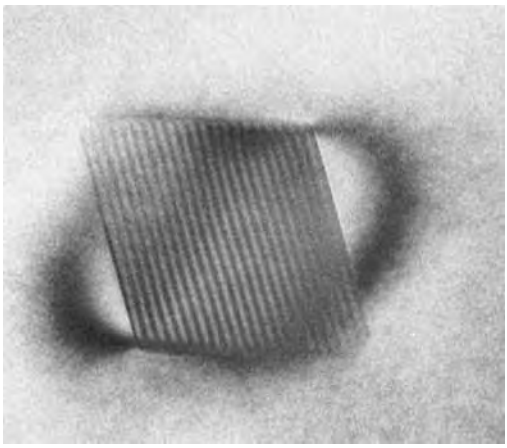


Figure 78. Moiré pattern formed at the interface between voidite and the diamond matrix. Note the extinction contours revealing the strain field. (Courtesy of G. Van Tendeloo.)

Applications of Moiré Patterns

An important application of parallel moiré fringes is the determination of the lattice parameter of one of the two components in a sandwich, the lattice parameter of the other being known. This can be of interest for the identification of plate-like coherent precipitates in a matrix with a known lattice parameter. Moiré fringes formed at the interface between voidite and diamond are shown in Fig. 78, and Fig. 79 shows the moiré fringes at the interface between silicon and silicon carbide (SiC) precipitate particles.

Moiré fringes have also been used as a tool in the study of dislocations. Ending moiré fringes reveal the emergence points of dislocations in one of the two components of the sandwich. The number, N, of supplementary half-fringes depends on the reflection used to produce the dislocation image; it is given by $N = \boldsymbol{g} \cdot \boldsymbol{b}$. This number is independent of the character of the dislocation. Supplementary half-fringes cannot therefore be interpreted as necessarily meaning that the corresponding dislocation has edge character. Partial dislocations bordering stacking faults are revealed by a 'fractional' number of

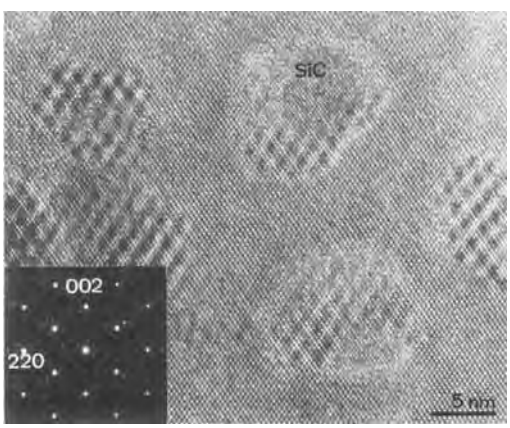


Figure 79. Moiré pattern at the interface between silicon and a silicon carbide precipitate. (Courtesy of A. De Veirman.)

supplementary half-fringes; that is, along the trace of the stacking fault the moiré fringes are shifted discontinuously over a fraction $\boldsymbol{g} \cdot \boldsymbol{b}$ such as $\frac{1}{3}$ or $\frac{2}{3}$ of the interfringe distance.

The moiré fringes are also shifted by a surface step in one of the components. The fringe shift is not only a function of the step height but also of the deviation parameter and hence of the specimen orientation.

If two or more diffraction vectors are active in both components of the sandwich a crossed grid of moiré fringes is formed, which has the rotation symmetry of the two films.

1.1.1.20 Applications of Diffraction Contrast

Diffraction contrast images do not reveal the crystal structure but they do image defects such as dislocations, planar interfaces (stacking faults, coherent domain boundaries, and out-of-phase boundaries), discommensurations, point defect clusters. Applications of diffraction contrast to specific classes of materials are described in Vol. II of this handbook. Here we discuss only a few characteristic examples illustrating different types of defect study. The interpretation of the images is only meaningful in terms of the underlying materials problem. Therefore, for each application we sketch as briefly as possible the framework in which the images acquire their significance and interpretation. The choice of the examples, which is admittedly subjective, is mainly motivated by the availability of suitable photographs.

The Fine Structure of Dislocations

Measuring the Stacking Fault Energy

In most materials the dislocations are not simple line defects but consist in fact of two or more partial dislocations connected by strips of stacking fault or of out-of-phase boundary. The simplest situation arises when glide takes place between two close-packed layers of identical 'spherical' atoms in an elemental face-centered-cubic crystal. The glide motion along the (111) plane in the $[\bar{1}10]$ direction then follows the valleys, that is it takes place in two steps, each performed by the motion of a partial dislocation, the first with a Burgers vector $\boldsymbol{b}_1 = \frac{1}{6}[\bar{2}11]$ and the second with a burgers vector $\boldsymbol{b}_2 = \frac{1}{6}[\bar{1}2\bar{1}]$, enclosing an angle of 60° and leading to a symmetry translation $\frac{1}{2}[\bar{1}10]$ along the (111) glide plane. Between the two partial dislocations a stacking fault ribbon with a displacement vector, equal to one of the Burgers vectors of the partial dislocations, is present [53].

The two partial dislocations repel one another, since their Burgers vectors enclose an acute angle. In an infinite solid this repulsion is proportional with $1/d$ (where d is the partial dislocation separation) and its magnitude is a function of the orientation of the partial dislocations that is a function of their character (screw or edge). The presence of the stacking fault ribbon causes an effective attractive force per unit length between the two dislocations, which is independent of their separation and numerically equal to the stacking fault energy, γ. An equilibrium separation is thus established. Assuming the repulsive force law to be known, it is then possible to deduce the stacking fault energy from the measured equilibrium separation of the

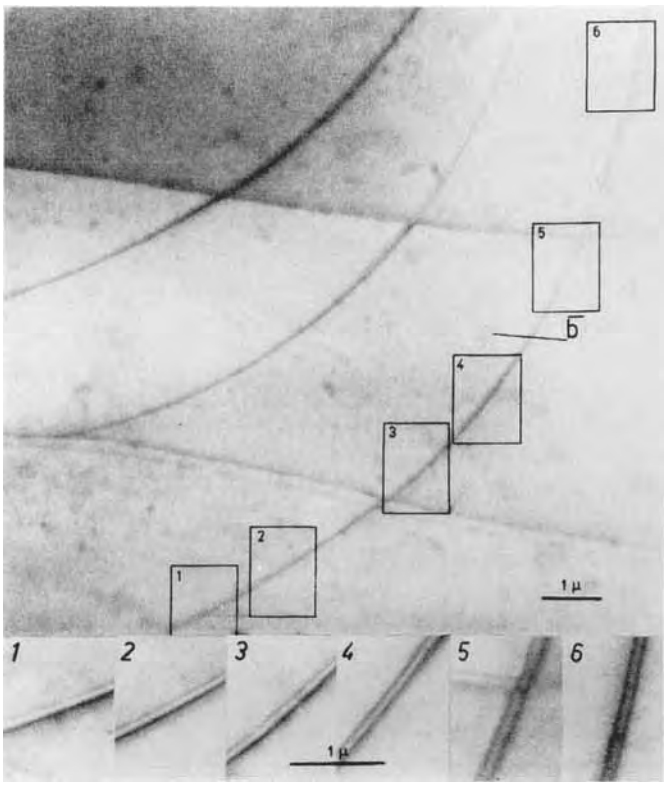


Figure 80. Curved dislocation ribbon in the (0001) plane of graphite. Several segments are reproduced as magnified insets. The direction ***b*** of the total Burgers vector, as determined by extinction experiments, is indicated. Note the systematic change in width with orientation [56].

partial dislocations. Dislocation ribbons are thus sensitive probes for measuring the stacking fault energy, a quantity which is difficult to access in any other direct way. The following relations apply in an infinite isotropic solid:

$$d = d_0\left[1 - \left(\frac{2\nu}{2-\nu}\right)\cos(2\phi)\right] \qquad (181)$$

with

$$d_0 = \frac{\mu b^2(2-\nu)}{8\pi\gamma(1-\nu)} \qquad (182)$$

where ϕ is the angle between the total Burgers vector and the ribbon direction; μ is the shear modulus and ν is Poisson's ratio.

The orientation dependence of the ribbon width can be verified on an image such as the one shown in Fig. 80, which represents a curved dislocation in a graphite foil. The Burgers vectors were determined using the method described in Sec. 1.1.1.18 of this Chapter. Plotting d as a function of $\cos(2\phi)$ the slope of the straight line so obtained gives the effective value of the Poisson ratio as well as the intercept d_0 to be used in the second relation, which then yields a value for the stacking fault energy (Fig. 81).

Using this method it is implicitly assumed that the repulsive force between dislocations is proportional to $1/d$, which is only the case in an infinite solid. In a thin foil the repulsive force between dislocations parallel to the foil surfaces decreases with decreasing distance to the specimen surfaces. This behavior can be observed, as

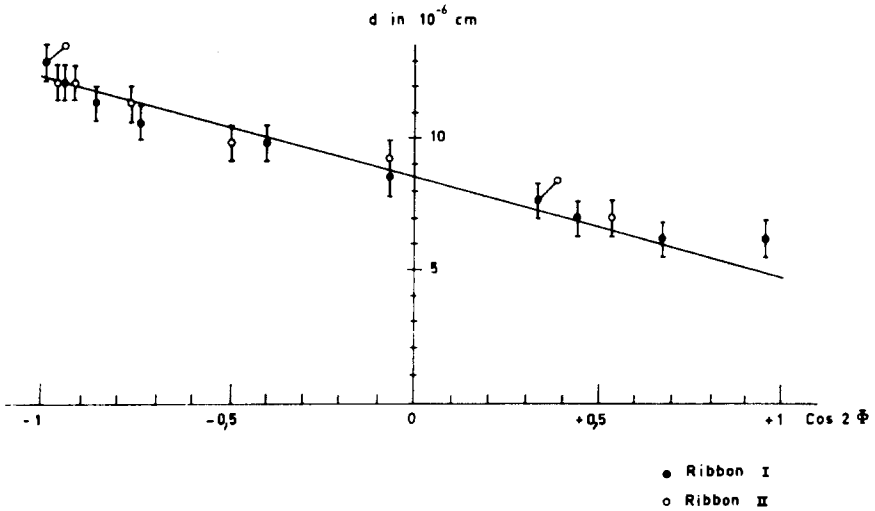


Figure 81. Plot of ribbon width d as a function of $\cos 2\phi$ [56].

shown in Fig. 82, where a ribbon gradually approaching the surface in a wedge-shaped lamella of tin disulphide, closes as it emerges at the surface [53, 54].

The energy of a dislocation ribbon depends on its distance to the surface. As a result, the shape of minimum energy of a dislocation ribbon crossing a surface step is not a straight line; 'refraction' of the ribbon as well as a change in width occur on passing underneath the surface step (Fig. 83). The *index of refraction* is the ratio of the total energies of the ribbon in

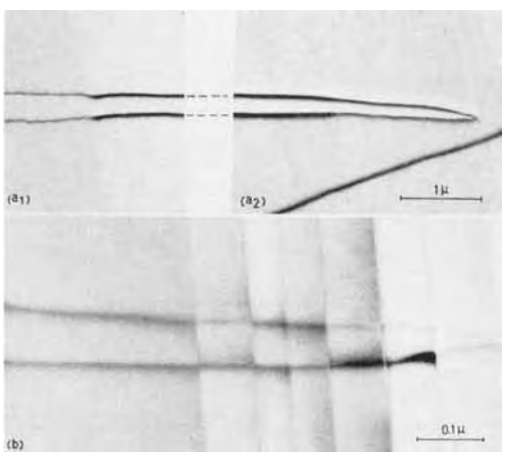


Figure 82. A wide ribbon in SnS_2 gradually approaching the surface. As the ribbon crosses surface steps it becomes discontinuously narrower. The ribbon closes where it emerges in the surface [54].

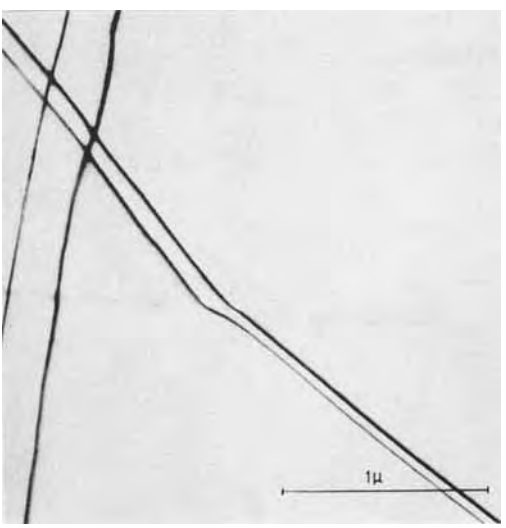


Figure 83. Dislocation ribbon in SnS_2. Refraction, accompanied by a width change, occurs on passing underneath a surface step [54].

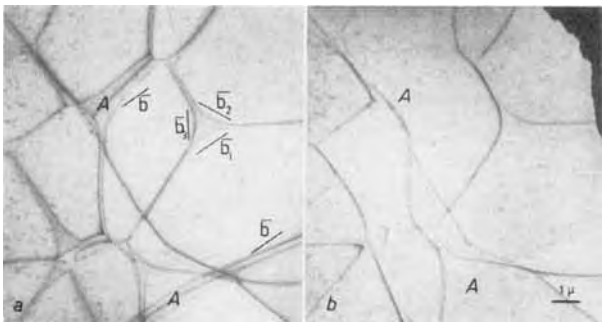


Figure 84. (a) Widely extended dislocation node of partial dislocations in graphite. At A a triple ribbon is present; the three partials have the same Burgers vector as follows from the contrast experiment in (b) where the three partials are simultaneously out of contrast. Nevertheless, the contrast at the three partial dislocations is different in (a) [56].

the two parts of the foil on either side of the surface step.

These images prove that such surface effects are not negligible. Therefore, when measuring stacking fault energies care should be taken to use foils of maximum thickness compatible with the visibility of the dislocations and, moreover, take the widest ribbon as the most representative one. The width of narrow stacking fault ribbons can best be determined by imaging in the weak-beam mode (see Sec. 1.1.1.17 of this Chapter).

Other geometrical configurations involving stacking faults can be used, such as the separation of partial dislocations in triple ribbons in graphite (Fig. 84a) and in close-packed structures, or the radius of curvature of single partial dislocations in a network of extended–contracted nodes (Fig. 84b). In the latter case one has, approximately,

$$\gamma = \frac{\mu b^2}{2R} \qquad (183)$$

where R is the radius of curvature and b is the Burgers vector of the partial dislocation. Isolated extended nodes such as the one shown in Fig. 85 (observed in AlN) are particularly suitable. More accurate relations are discussed by Nabarro [53] and Siems et al. [55].

Multiribbons

Ordering in alloys based on close-packed structures leads to long symmetry translations along the glide directions in the close-packed glide planes. This results in ribbons consisting of several partial dislocations, separated either by stacking faults or by out-of-phase boundaries. The equilibrium separation of superdislocations (i.e., perfect dislocations with respect to the basic lattice, but partial dislocations with respect to the ordered structure) can be used to derive values of the antiphase boundary energy in the same way as described above for stacking faults. In Ni_4Mo, as many as ten partial dislocations are connected by faults and antiphase boundaries [53].

The dislocations involved in glide between the close-packed layers of anions (X) in layered ionic sandwich crystals of CdI_2 of the type AX_2 (XAXXAX...) or AX_3 are of particular interest. The glide motion takes place between the two weakly van der Waals bonded close-packed anion layers. Dislocations can thus dissociate into two or more Shockley partial dislocations. Although in the close-packed layers, between which glide takes place, all X atoms seem equivalent, the A cations in the adjacent central layers of the

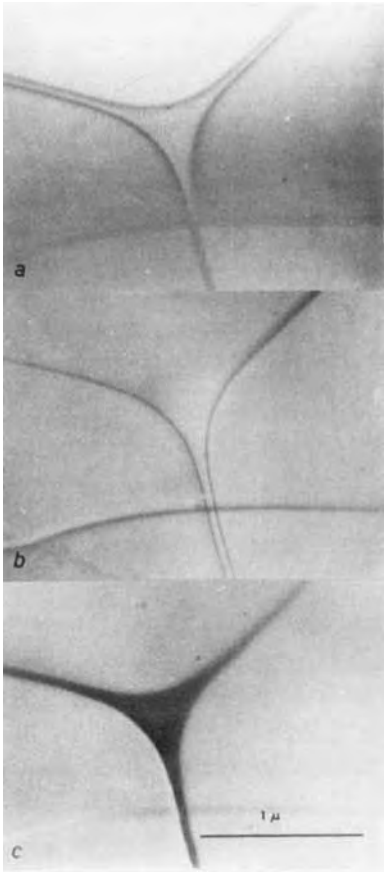


Figure 85. Isolated extended node in the (0001) plan of AlN. (Courtesy of P. Delavignette.)

sandwiches may form configurations which impose a large unit mesh in the glide plane, either because not all octahedral cations sites are occupied (e.g., in $CrCl_3$ and $CrBr_3$) or because the cations form metal–metal bonded clusters leading to a slight deformation of the close-packed layers (e.g., in $NbTe_2$ and $TaTe_2$) and the creation of a superperiod.

In the chromium trihalides [53, 56] multiribbons containing either four or six partial dislocations are observed. Assuming glide along the close-packed anion layers to take place by the propagation of Shockley partial dislocations, two types of stacking fault ribbons can be distinguished:

(i) faults violating only the chromium stacking (i.e., involving only third neighbors), and
(ii) faults violating the stacking of the chromium ions as well as that of the anions (i.e., involving next-nearest neighbors).

Intuitively, it is clear that the type (ii) faults will have a larger energy than those of type (i). The six-fold ribbons correspond to a 'straight' zigzag glide path along the close-packed directions in the (0001) glide plane of the anion sublattice; they contain the two types of fault in an alternating fashion, the outer ribbons corresponding to high energy faults. Diffraction contrast images of such ribbons are shown in Fig. 86. The outer ribbons are clearly the narrowest ones. The structure of the four-fold ribbons can similarly be related to the structure.

Also in $NbTe_2$ [57, 58], which has a deformed CdI_2 structure, six-fold ribbons

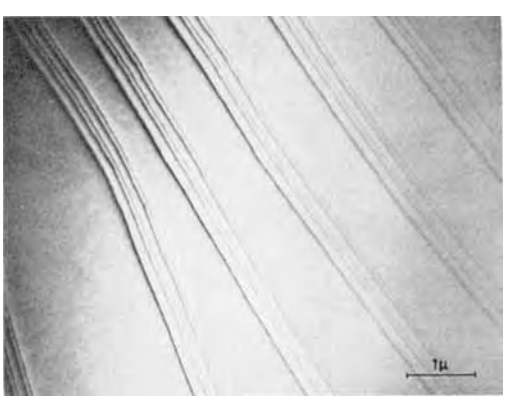


Figure 86. Six-fold ribbons of partial dislocations in the (0001) plane of $CrCl_3$. The Burgers vectors of the partial dislocations form a zigzag glide path [53, 56].

occur. In this structure the niobium ions form clusters of three parallel close-packed rows, having a somewhat smaller separation than in the ideal hexagonal structure which probably occurs only in the temperature range in which the crystal is grown. The resulting structure then becomes monoclinic on cooling. The unit mesh in the glide plane is a centered rectangle, which can adopt three different but equally probable orientations differing by 60°. As a consequence, the room temperature structure is fragmented into domains corresponding to the three possible orientations of the clustered niobium rows. The monoclinic symmetry causes the glide paths along the three close-packed directions within the same domain to become non-equivalent.

The zigzag glide paths in the direction enclosing an angle of 30° with the long side of the rectangular mesh consist of six partial dislocations, whereas the glide path along the other close-packed direction (i.e., along the short side of the rectangle) repeats after two partial dislocations. The Burgers vector is conserved all the way along the dislocation lines. Hence when a six-fold ribbon passes through a domain wall the glide path changes its orientation relative to the underlying structure. A six-fold ribbon in one domain is thus transformed into three separate two-fold ribbons in the adjacent domain. Whereas in the six-fold ribbon the six partial dislocations are held together by stacking faults, this is no longer the case with the three two-fold ribbons which repel one another and hence develop 'bulges'. The image shown in Fig. 87 illustrates the behavior of a six-fold ribbon intersecting a set of parallel domain boundaries in $NbTe_2$ [58].

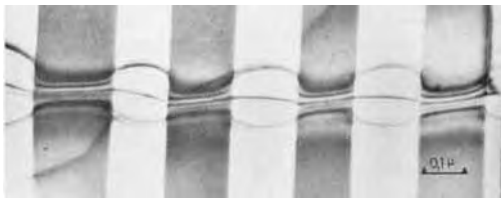


Figure 87. Six-fold ribbon of partial dislocations in $NbTe_2$ intersecting domain boundaries along which the underlying structure changes by 60° in orientation. In half of the domains the six-fold ribbons separate into three two-fold ribbons which form bulges as a result of repulsive forces [58].

Plastic Deformation: Glide Dislocations

Plastic deformation is a subject which was studied intensely in its early stages by means of diffraction contrast. High voltage (about 1000 kV) electron microscopy has been of considerable interest in this respect because it can be used to study thicker foils, which are more representative of a bulk material than the thin foils required at 100 kV. Figure 88 shows a procession of glide dislocations in face-centered cubic stainless steel, confined to their (111) glide plane, as observed in high voltage electron microscopy. The strictly planar arrangement implies that the dislocations are dissociated and that, for this reason, the cross-glide is a difficult process. The dissociation is too small to be directly observable at this resolution, but it has been found from other images that the stacking fault energy is rather low in stainless steel. Note the periodic contrast of the dislocations in the vicinity of their emergence points in the foil surfaces and the absence of such contrast in the central part of the foil. Figure 89 is an image of a low stacking fault energy alloy (Cu-Ga); the dissociation is clearly visible and stacking fault fringes can be observed between the partial dislocations.

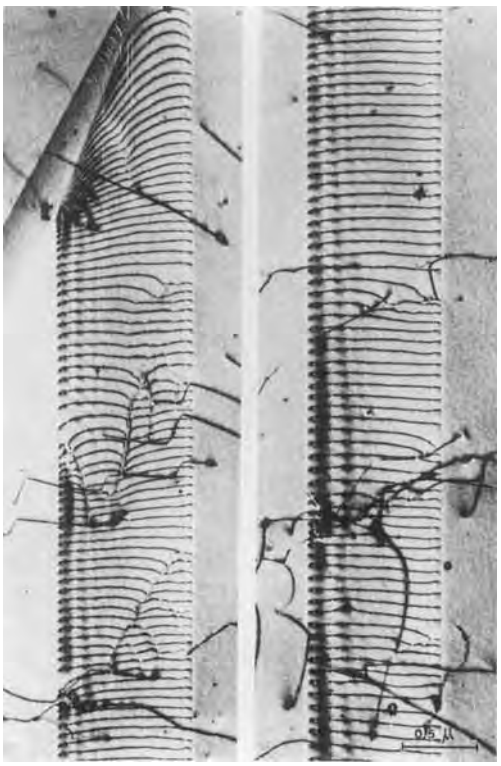


Figure 88. High voltage electron micrograph showing a procession of dislocations confined to a glide plane in stainless steel. Note the wavy contrast close to the surfaces and its absence in the central part of the foil.

Figure 89. Ribbons dissociated in Shockley partials observed in a Cu-Ga alloy. (Courtesy of A. Art.)

Figure 90 shows a network of intersecting glide dislocations confined to the (111) glide plane in a face-centered cubic copper alloy (Cu-Ga) with a low stacking fault energy. One set of dislocation nodes is dissociated and gives rise to the dark triangular areas; the other set is contracted. Such nodes allow one to deduce the stacking fault energy from the curvature of the partial dislocations forming the extended nodes.

Figure 91 shows glide dislocations in the layer plane (001) of $NbTe_2$, which is parallel to the foil plane, the specimen having been obtained by cleavage [58]. In every other domain the dislocation multiribbons consist of six partial dislocations; in the remaining domains the dislocations are simple ribbons, as described above. The image illustrates the strong interactions of the glide dislocations with the domain walls. On entering a domain in which the six-fold ribbons would have to be formed, the single ribbons line up with the domain wall, in this way minimizing the generation of stacking faults. This leads to an effective interaction between the dislocation ribbons and the domain walls.

The Structure of Subgrain Boundaries

Small-angle grain boundaries can be described in terms of arrays of dislocation

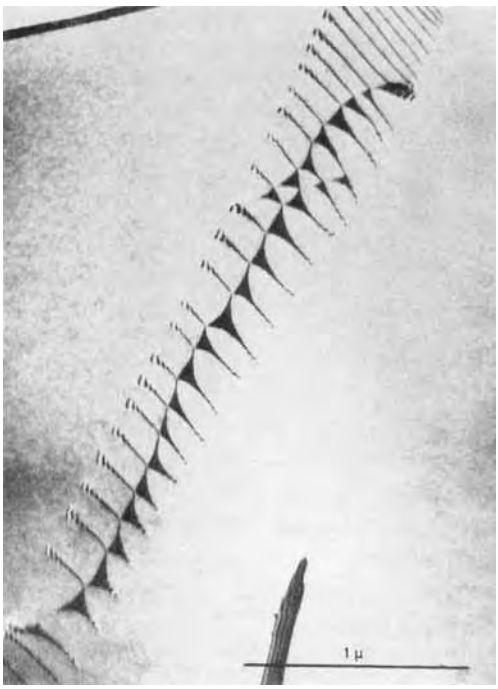


Figure 90. Network of dissociated dislocations in a Cu-Ga alloy with a low stacking fault energy. (Courtesy of A. Art.)

lines. Diffraction contrast electron microscopy has contributed significantly to firmly establishing dislocation models for such boundaries.

A general subgrain boundary is characterized by five parameters describing its geometry: the rotation axis, the rotation angle, and the normal to the contact plane. These parameters can be determined by the combined use of the spot diffraction pattern and of the Kikuchi line pattern. The diffraction contrast image then allows one to visualize the geometry of the dislocation lines and, using the extinction criterion, to determine their Burgers vectors. If the rotation axis is parallel to the contact plane the boundary is a tilt boundary and the dislocation configuration consists of parallel lines. If, on the other hand, the rotation axis is perpendicular to the contact plane the boundary consists of a network of intersecting dislocations. Depending on the symmetry of the contact plane this network may ideally consist of square meshes in (100) planes or of hexagonal meshes in (111) planes.

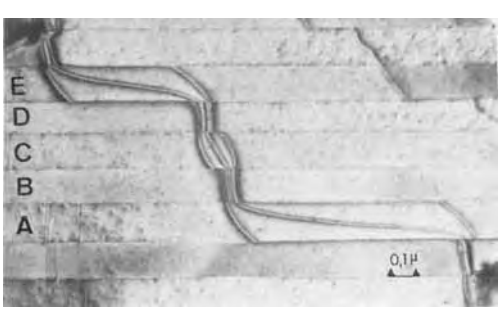


Figure 91. Glide dislocations in the layer plane of $NbTe_2$. Note the interaction between dislocations and twin domain walls [58].

Figure 92 shows two tilt boundaries in body-centered niobium consisting of parallel dislocation lines. Some of the dislocations are decorated by small precipitate particles.

The image of the twist boundaries in platinum [59] shown in Fig. 93a illustrates a hexagonal network containing three intersecting families of dislocations with Burgers vectors enclosing angles of 120°. Figures 93b and 93c represent square networks, consisting of dislocations with mutually perpendicular Burgers vectors.

Figure 94 shows a well-developed hexagonal network of undissociated dislocations in the (0001) plane of hexagonal zinc [60]. All dislocations are mobile in the plane of the boundary. The right part of the boundary moved along the (0001) glide plane during the exposure, leading to blurring of the image. In the

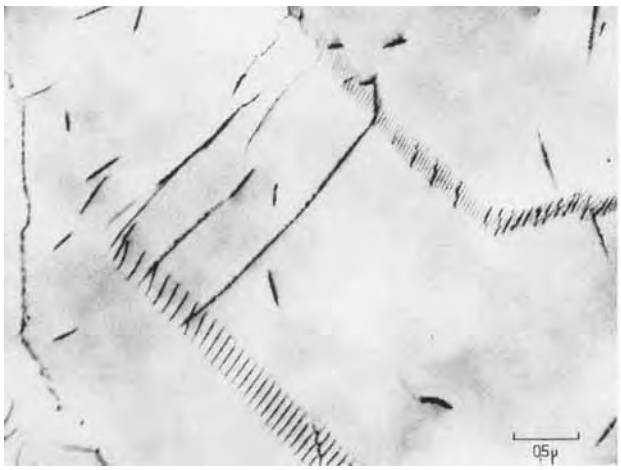


Figure 92. Tilt boundary consisting of sets of parallel dislocations in niobium. Some of the dislocations are decorated by small particles. (Courtesy of A. Fourdeux and A. Berghezan.)

left part of the image some of the dislocations, leaving the network and terminating in the foil surfaces, have become sessile, and they thus pinned the network in that part.

Figure 21 shows a hexagonal network of widely extended dislocations in the basal plane of graphite [56, 61]. The network is, in fact, a glissile twist boundary. From the curvature of the partial dislocations in the extended nodes one can deduce the stacking fault energy.

Point Defect Clusters

Vacancies in quenched metals form disc-shaped agglomerates in (111) face-centered-cubic or (0001) hexagonal-close-packed layers, limited by Frank-type dislocation loops. If the stacking fault energy is large enough the loop is 'unfaulted', since energy is gained by nucleating a Shockley partial dislocation and sweeping the loop, transforming the sessile Frank loop into a perfect glissile loop. Such unfaulted loops

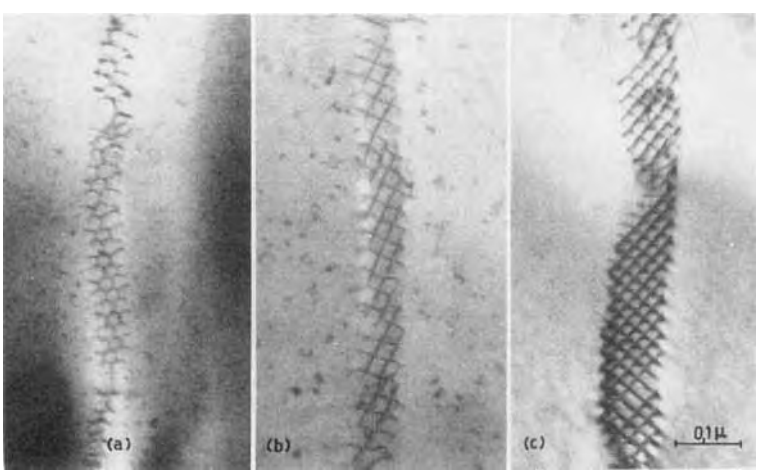


Figure 93. Twist boundaries in platinum [59]: (a) hexagonal network; (b, c) square networks. (Courtesy of E. Ruedl).

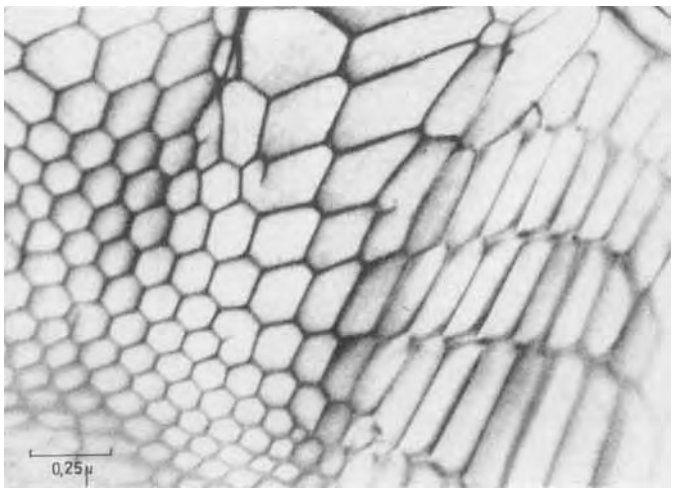


Figure 94. Hexagonal network of undissociated dislocations in the (0001) plane of zinc [60].

in quenched aluminum are shown in Fig. 95 [62, 63].

If the stacking fault energy is small enough, which is true in, for instance, gold and Ni-Co alloys, the Frank loop is transformed into a stacking fault tetrahedron consisting of four intersecting triangular stacking faults in {111} planes, limited along their intersection lines by edge-type stair-rod dislocations with a Burgers vector of the type $\frac{1}{6}[110]$. For intermediate values of the stacking fault energy the Frank loops may remain faulted. Figure 96 shows stacking fault tetrahedra in gold imaged in diffraction contrast [62, 64].

Faulted Frank loops in silicon are shown imaged in diffraction contrast in

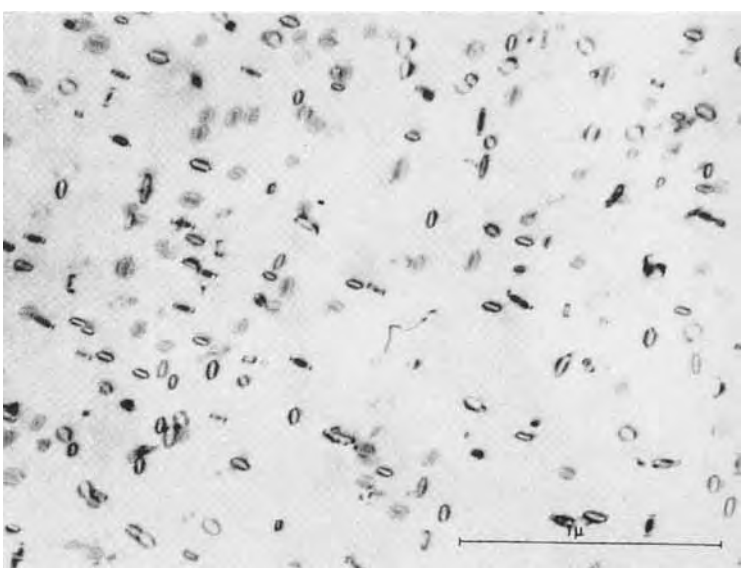


Figure 95. Unfaulted dislocation loops in quenched aluminum [62, 63].

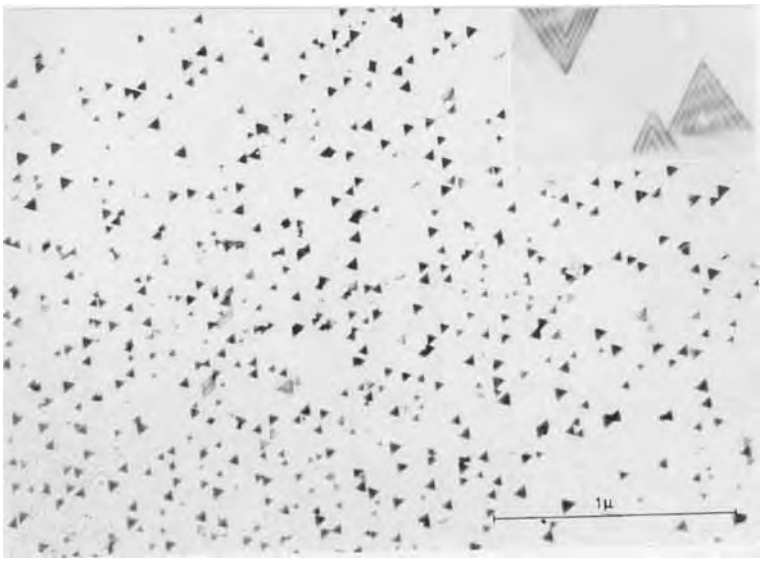


Figure 96. Diffraction contrast image of stacking fault tetrahedra in quenched gold. The inset shows a magnified image [62, 64].

Fig. 97. The presence of the stacking fault causes contrast inside the loop. Figure 98 shows a contrast experiment on an extrinsic Frank-type dislocation loop in silicon. Note the presence of a line of no contrast perpendicular to the active $\boldsymbol{g}$ vector and the deformation of the extinction contours where they cross the dislocation loop. For $\boldsymbol{g} = [1\bar{1}1]$ the loop exhibits stacking fault contrast, as do the loops in Fig. 97.

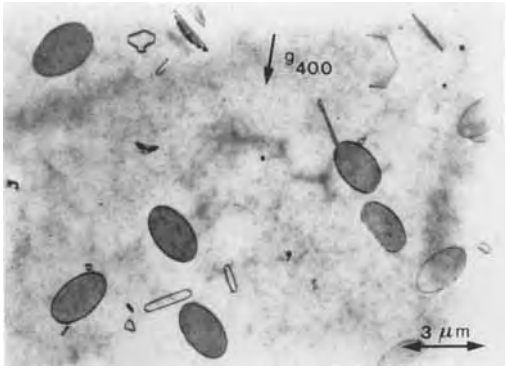


Figure 97. Faulted Frank loops in silicon due to interstitials. The loops exhibit stacking fault contrast. (Courtesy of H. Bender.)

Planar Interfaces

It is well known that two simple types of stacking fault can occur in the face-centered-cubic structure. The *intrinsic fault*, formed either by the extraction of a layer or by glide of a Shockley partial, is represented by the stacking symbol

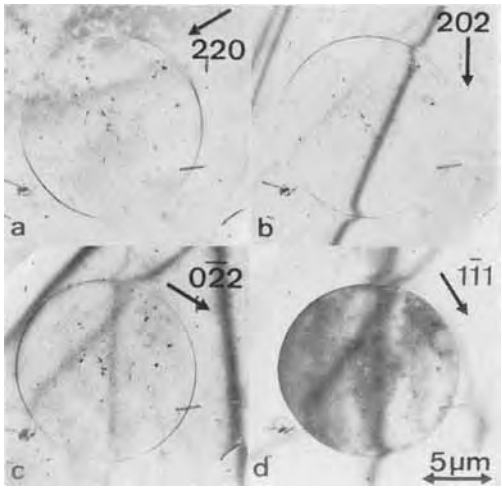


Figure 98. Contrast experiment on a faulted loop in silicon.

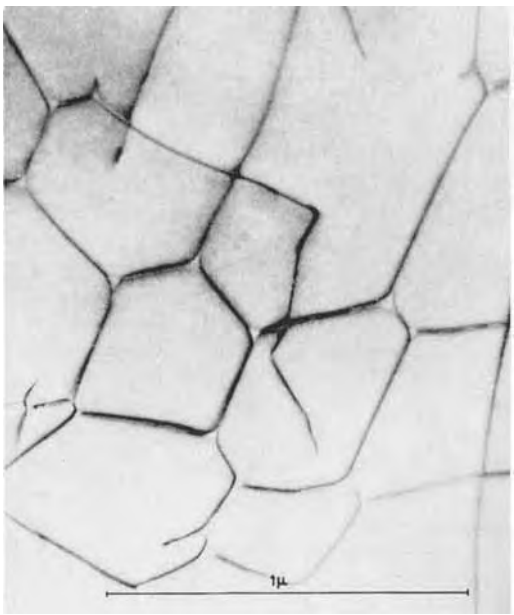


Figure 99. Network of extended dislocations in silicon; all nodes are dissociated. (Courtesy of E. Aerts.)

abcabcacabc.... The *extrinsic fault*, formed for instance by the precipitation of interstitials in a Frank loop, corresponds to the stacking symbol abcabacabc.... The two types of fault have comparable energies in certain materials. In a network of dissociated dislocations all nodes are then dissociated; this is, for instance, the case in silicon (Fig. 99) and in certain alloys (e.g., Ag-Sn). The two kinds of fault have opposite displacement vectors of the type $\boldsymbol{a}/3$ [111] and can thus be distinguished by the characteristic fringe pattern which they produce when situated in inclined planes (see Sec. 1.1.1.13 and Sec. 1.1.1.20 of this Chapter) [65–67].

Domain Structures

Phase transformations are usually accompanied by a decrease in symmetry with decreasing temperature. As a result, a single crystal of a higher symmetric phase becomes fragmented into domains of which the structures are related by the symmetry elements lost in the transition to the lower symmetry phase [68]. The lost rotation symmetry elements give rise to orientation variants of the low temperature phase of which the number is given by the ratio of the order of the point group of the high temperature phase and the order of the point group of the low temperature phase. The loss of translation symmetry gives rise to translation variants related by displacement vectors given by the lost lattice translations. Their number is determined by the ratio of the volumes of the primitive unit cells of the low and high temperature phases.

Orientation variants are separated by domain boundaries, whereas translation variants are separated by out-of-phase boundaries. The orientation of the domain boundaries is determined by the requirement that the strain energy should be a minimum. This will be the case for strain-free interfaces. As a result, the orientation of certain interfaces (W) follows entirely from symmetry, whereas others (W′) have orientations which depend on the lattice parameters of the two phases involved, at the transition temperature [69–71].

For example, in the α–β transition of quartz, referred to above, the α phase has the point group 32 (order 6) and the β phase has the point group 622 (order 12). The number of orientation variants in the α phase is thus $12/6 = 2$ (α_1 and α_2; Dauphiné twins) and they are related by the lost 180° rotation about the three-fold axis. There is no change in translation symmetry. Images of domain fragmented α phase are shown in Fig. 100. In the case

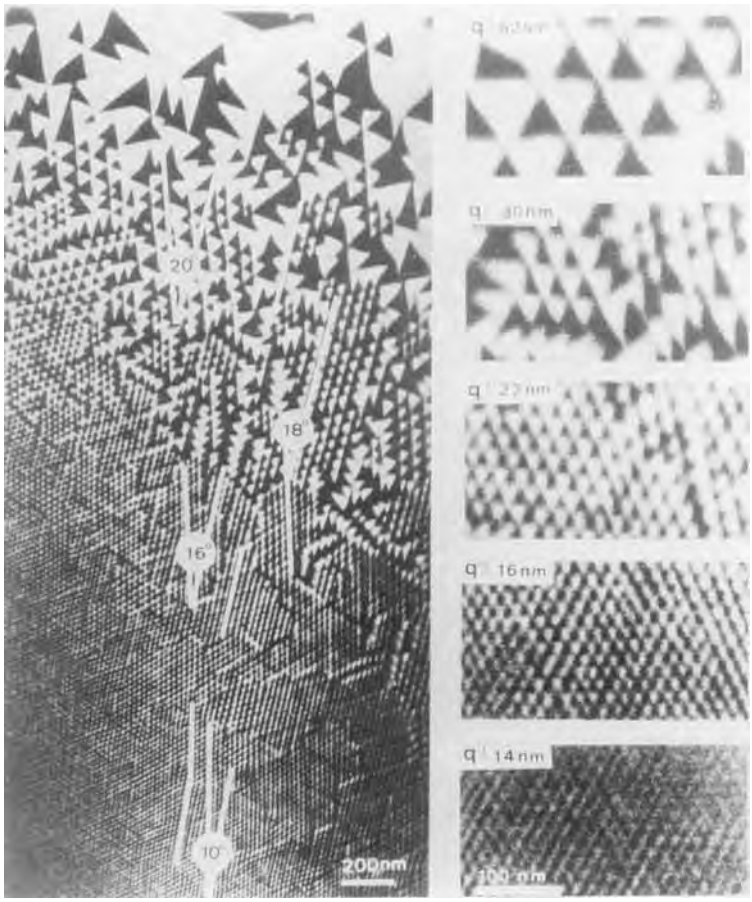


Figure 100. Domain fragmentation in quartz as a function of temperature. A temperature gradient is present across the specimen. At the highest temperature the incommensurately modulated phase is observed [28].

of quartz the situation is actually somewhat more complicated by the occurrence of an intermediate incommensurate phase between the α and β phases and which is only stable within a narrow temperature range ($\approx$1.5 K). This phase was discovered using diffraction contrast electron microscopy [30]. It consists of a regular texture of triangular prisms parallel to the c axis, of α_1 and α_2 structure. The size of the triangular prisms decreases with increasing temperature in the vicinity of the transition temperature (Fig. 100).

Quite striking domain structures were studied using diffraction contrast in the monoclinic room temperature phase of ferroelastic lead orthovanadate $[Pb_3(VO_4)_2]$ [70, 71]. The structure is rhombohedral at high temperature (γ phase), but on cooling it transforms at 120°C into a monoclinic structure (β phase) which is stable at room temperature. The rhombohedral parent phase is fragmented into domain patterns which minimize the strain energy. They consist of combinations of completely symmetry determined walls (W) and walls (W′), the strain-free orientation of which depends on the lattice parameters below and above the $\gamma \leftrightarrow \beta$ transition temperature

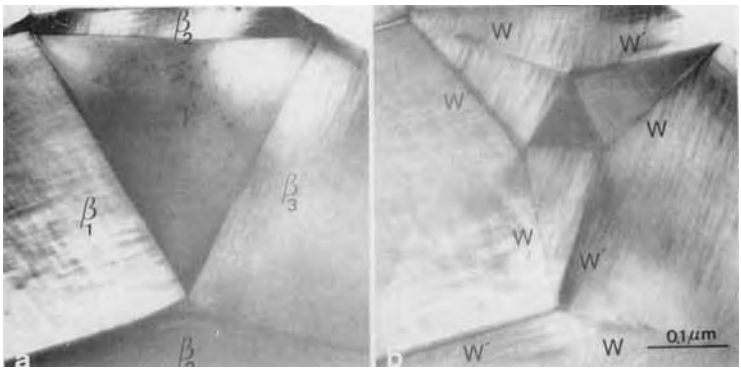


Figure 101. Domain pattern in lead orthovanadate $[Pb_3(VO_4)_2]$ resulting from the $\gamma \rightarrow \beta$ phase transition. The central triangle of the star pattern is still in the γ phase. The two images refer to the same area; in (b) the temperature was somewhat lower than that in (a) [71].

(i.e., on the spontaneous strain tensor). The most striking configuration is the pattern shown in Fig. 101; it contains three concentric 'stars' of decreasing size. The pattern in Fig. 101a shows a central triangle of metastable γ phase surrounded by areas consisting of three different variants of the β phase. On cooling this triangle transforms further 'in situ' into the configuration shown in Fig. 101b; the γ triangle becomes smaller and is rotated by 180°. Similar patterns occur in other domain textures resulting from a phase transformation between the parent and product phases belonging to the same point groups as γ and β lead orthovanadate, respectively.

The compound $YBa_2Cu_3O_7$ is tetragonal at high temperature where the $\cdots$O–Cu–O–Cu$\cdots$ chains in the CuO layers are disordered. Below the transition temperature, which depends on the oxygen content, the chains order in any given area along one out of two mutually perpendicular, equally probable, orientations, which then becomes the b_0 direction of the orthorhombic structure. The disorder–order transition thus produces two structural variants with their b_0 axes roughly perpendicular and which are twin related by a mirror operation with respect to (110) or $(1\bar{1}0)$. These two orientation variants are revealed, using different imaging modes, in Fig. 102 [72].

The ordering of magnetic moments below the Néel temperature in antiferromagnetic materials is usually accompanied by a structural phase transition. This leads to the formation of an antiferromagnetic domain structure of which the domain walls coincide with those due to the structural phase transition [73]. Such a combined transition occurs, for instance, in NiO, which has the sodium chloride structure above the Néel point (525 K). Below this temperature the nickel spins order in such a way that the spins in one of the families of (111) planes order ferromagnetically, the spin direction being parallel to these (111) planes, whereas successive (111) sheets contain oppositely oriented spins. As a consequence of magnetostriction, the structure contracts along the ⟨111⟩ direction perpendicular to these sheets, and the lattice becomes rhombohedral ($\alpha = 90° 4'$). The rhombohedral structures in adjacent antiferromagnetic domains contract along different ⟨111⟩ directions, and as a result such domains are separated by coherent twin boundaries with a very small twinning vector, which are imaged as δ-fringe patterns. Two such

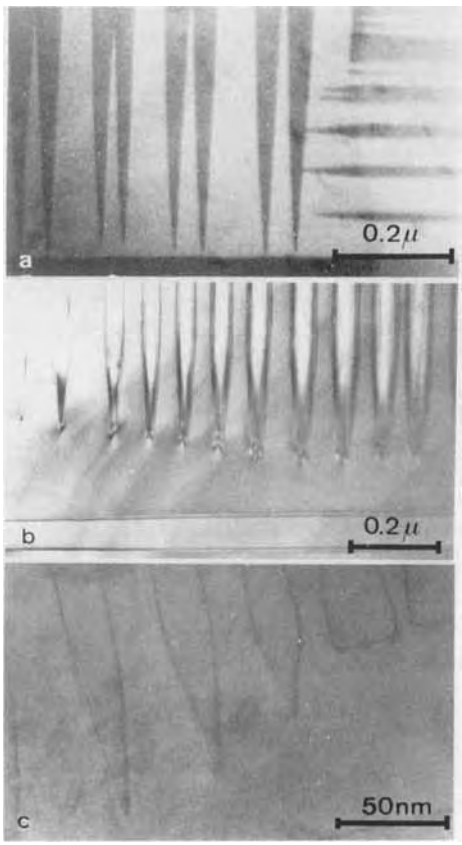


Figure 102. Orthorhombic twins in $YBa_2Cu_3O_{7-\delta}$ revealed using three different contrast modes: (a) domain contrast; (b) interface contrast; (c) high resolution imaging. Note the strain at the twin tips in (b) [72].

domain walls are shown in Fig. 103 which is a bright-field image of two parallel domain walls for which the δ values are opposite in sign. This is reflected in the opposite nature of the edge fringes for the two boundaries.

The Structure of Ordered Alloys

Observations of 'discommensurations' and of 'discommensuration nodes' [74] were performed on the alloy $Ni_{3+}Mo$ [75], using diffraction contrast, at a time when the term 'discommensuration' had not yet been introduced. The interfaces shown in Fig. 104 were described as 'out-of-phase boundaries' with a displacement vector equal to one-quarter of a lattice vector. Although in alloys there is no essential difference between out-of-phase boundaries and 'discommensuration walls', the

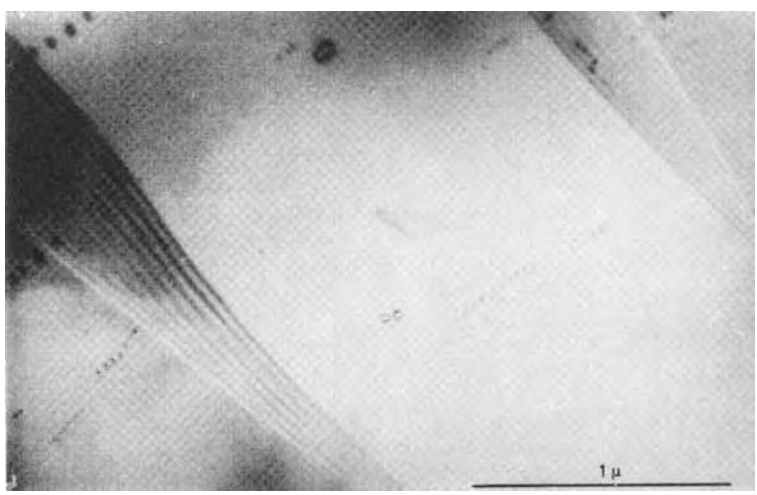


Figure 103. Two domain walls with opposite δ in antiferromagnetic nickel oxide; they are imaged as δ-fringe patterns [73].

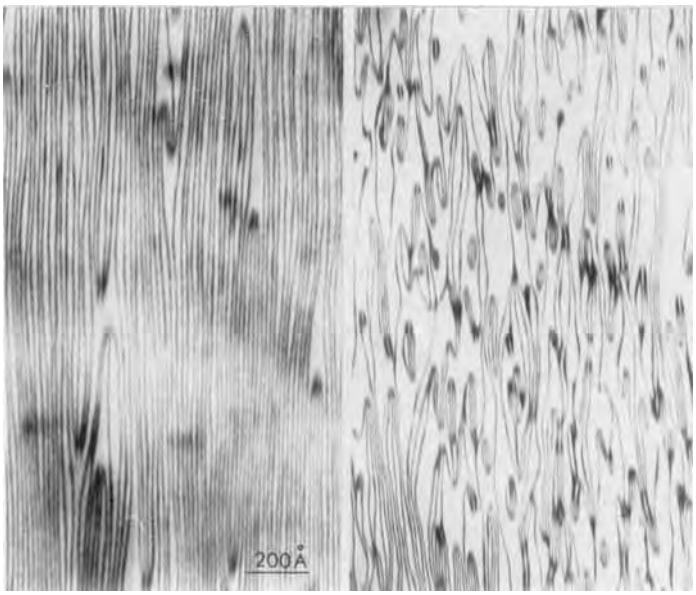


Figure 104. Four-fold discommensuration nodes in Ni_3Mo, revealed by diffraction contrast [75].

defects shown in Fig. 104 would, at present, presumably be termed 'discommensurations' by most authors.

Conservative antiphase boundaries in the alloy Cu_3Pd with $L1_2$ structure revealed by diffraction contrast are shown in Fig. 105. These boundaries represent the first stage in the formation of a one-dimensional long period antiphase boundary structure from the disordered phase. A number of nonconservative antiphase boundaries become unstable and start 'meandering', forming parallel sets of conservative antiphase boundaries [76, 77].

Minerals

Anorthite ($CaAl_2Si_2O_8$) is a complicated silicate which has a primitive triclinic Bravais lattice (space group $P\bar{1}$) at room temperature. Above $T_c = 514\,K$ the same unit cell becomes body centered ($I\bar{1}$). This can be concluded from the diffraction pattern since the spots of the type $h+k+l = \text{Odd}$ gradually disappear above T_c. On cooling the crystal from the high temperature phase to room temperature, it breaks up into two translation variants separated by very 'ragged' antiphase boundaries with a $\frac{1}{2}[111]$ displacement vector. No orientation variants are formed. The domain boundaries are

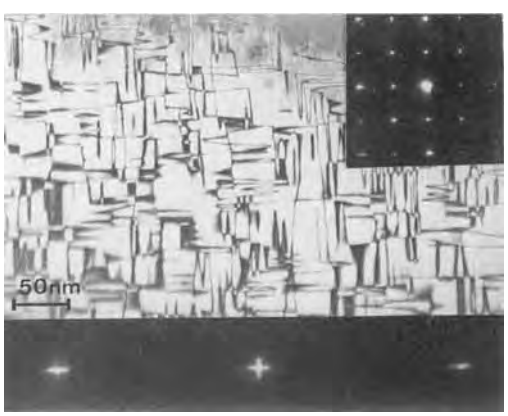


Figure 105. Diffraction contrast image of the first stage in the formation of a one-dimensional long period structure in Cu_3Pd. Note the 'meandering' of the antiphase boundaries [77].

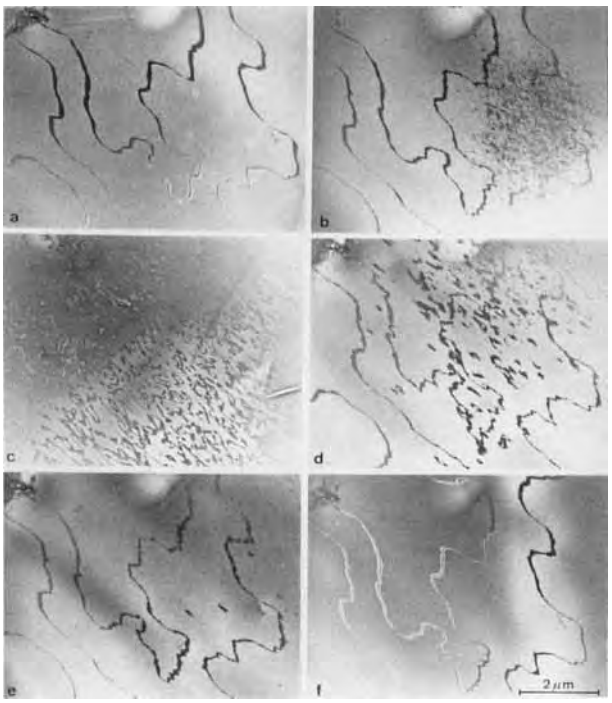


Figure 106. Evolution of antiphase boundaries in anorthite ($CaAl_2Si_2O_8$) during a heating–cooling cycle from room temperature up to above 514 K. All images refer to the same area. Note the memory effect [78].

revealed by diffraction contrast dark field imaging in reflections for which $h + k + l = \text{Odd}$. On heating above 514 K the boundaries disappear, but on cooling they reappear at exactly the same place and with the same shape as before (i.e., there is a pronounced memory effect, presumably due to impurity pinning). This is illustrated by the heating–cooling cycle shown in Fig. 106; the corresponding diffraction patterns along [101] are shown in Fig. 107 [78].

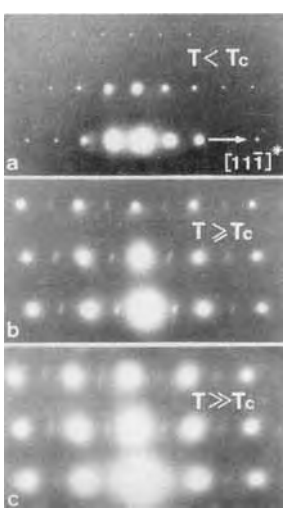


Figure 107. Evolution of the diffraction pattern during the same heating–cooling cycle as in Fig. 106: (a) $T < T_c$; (b) $T \geqslant T_c$; (c) $T \gg T_c$.

Fabrication-induced Defects in Semiconductors

Semiconductor single-crystal 'chips' often undergo a long sequence of fabrication steps (thermal treatment, oxidation, etching, etc.) some of which can be accompanied by a deterioration of the crystal's physical properties and hence affect the performance of the final device. The

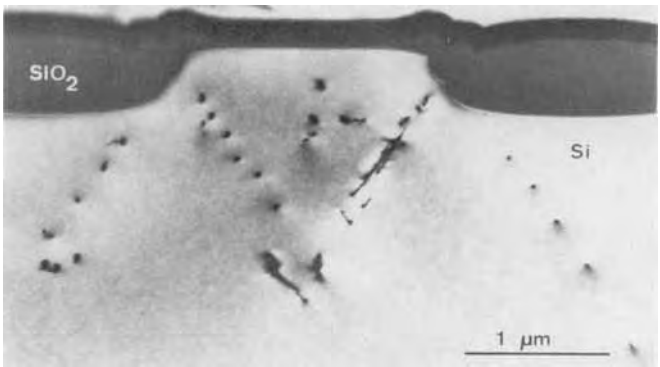


Figure 108. TEM image of a cross-section of a field-effect device. Dislocations are emitted from the edges of the constriction in the silicon oxide layer; the dislocations are seen end-on [79].

micro-miniaturization of electron devices makes detailed control of the crystal perfection strongly dependent on electron microscope techniques; both on high resolution images of cross-section specimens of devices, and on high voltage electron microscopy for the study of 'thick' specimens at low resolution and small magnification.

Figure 108 shows a transmission electron microscopy (TEM) image of processions of dislocations observed end-on in a cross-sectional view of a field-effect device. At the edge of the constriction in the silicon oxide layer sources have generated dislocations along the glide planes of maximum resolved shear stress, in order to relieve the stresses generated by the oxidation process. The dislocations apparently form 'inverse' pile-ups, their spacing being smallest close to the source [79].

Figure 109 shows 'finger'-shaped gate areas formed in a field oxide layer on a silicon chip. They have similarly generated stresses (see above) which are relieved by dislocation generation. In this case the dislocations are imaged in a plane view.

Oxide or other precipitate particles may put the surrounding silicon matrix under a compressive stress. This stress is often large enough to give rise to 'prismatic punching' whereby discs of self-interstitials surrounded by a loop of perfect dislocation are emitted. Such loops are glissite on a cylindrical surface of which the cross-section is determined by the precipitate's shape and the direction of the generators by the Burgers vector of the dislocations (i.e., $\frac{1}{2}\langle 110\rangle$) (Fig. 110).

Interfacial dislocations are often formed at the interface between the silicon substrate and metallic layers used as

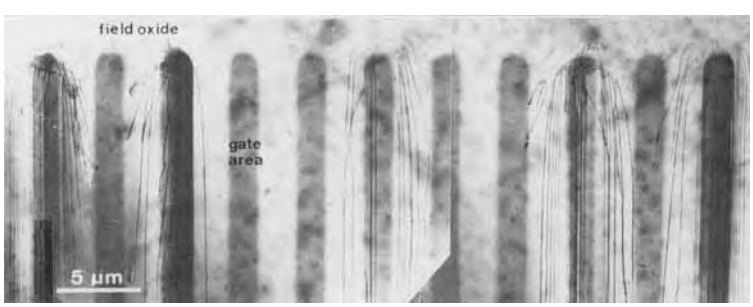


Figure 109. Finger-shaped gate areas in a field oxide. Dislocations are generated along the edges; they are observed in a plane view. (Courtesy of Vanhellemont.)

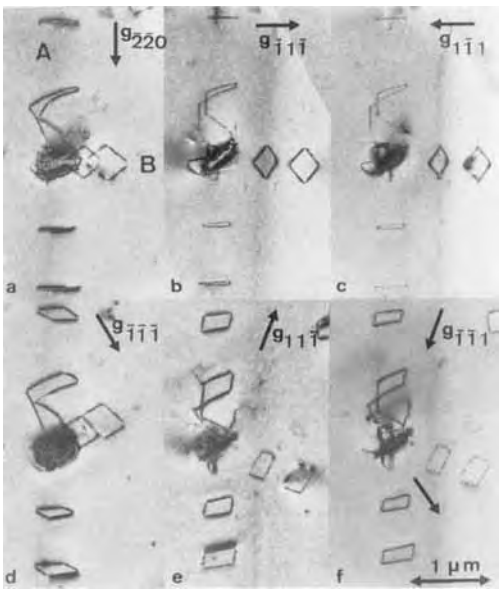


Figure 110. Prismatic punching around a precipitate particle in a silicon matrix imaged by six different diffraction vectors. (Courtesy of H. Bender.)

electrical contacts. Figure 111 shows the networks of misfit dislocations between silicon and aluminum; the image was obtained in a plane view using diffraction contrast.

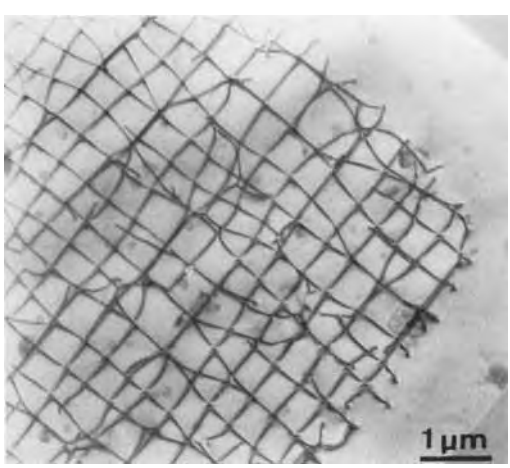


Figure 111. Network of interfacial dislocations in the contact plane between silicon and aluminum.

Various Applications

In Situ Studies

The availability of cooling and heating specimen holders allows in situ study of the phenomena accompanying phase transitions. When going through a disorder–order transition, different phases of the domain fragmentation can be followed. The creation and elimination of discommensuration walls is directly observable in dark field images made in clusters of incommensurate reflections [80, 81].

When performing such observations one should be aware of the effect of the electron beam on the specimen, which results in an increase in temperature (depending on the thermal conductivity of the foil) and may also cause some radiation effects (which may interfere with the transition).

Radiation Damage

Electron microscopy, in particular high voltage electron microscopy, has been used extensively to study in situ radiation effects as well as postirradiation defect configurations. The point defects, precipitates, and small dislocation loops can be characterized using the methods described above (see Sec. 1.1.1.18 of this Chapter).

Radiation Ordering

Some surprising results were obtained from in situ studies of ordering alloys that exhibit a short-range order state (e.g., Ni_4Mo). When irradiated with 1 MV electrons at low temperature, ordered Ni_4Mo becomes completely disordered. When irradiating in a temperature range below, but close to, the

order–disorder transition temperature, the irradiation causes the alloy to assume order up to a certain degree. The order parameters can be determined by following the evolution of the intensity of the order diffraction spots. These phenomena result from the competition between the ordering effect due to radiation-enhanced diffusion at the irradiation temperature and the disordering effect of the irradiation as a result of atomic collisions. Over a certain temperature range the short-range order state is produced by irradiation. Certain alloy phases which could not be ordered by thermal treatment (e.g., Ni_4W), were found to assume order under electron irradiation.

Determination of the Type of Stacking Fault

Close-packed layers of atoms can be stacked in an infinite number of ways, all of which have nearly the same free energy. Two essentially different types of stacking mode are usually distinguished: the face-centered-cubic stacking mode (...ABCABC...) and the hexagonal-close-packed mode (...ABAB..., or ...ACAC..., or ...BCBC...).

In face-centered stacking two different types of stacking fault are often distinguished. If a single atomic layer is extracted and the gap so created is closed by a displacement over a vector $\boldsymbol{R}_0 = \frac{1}{3}[111]$, the resulting sequence is

$$\ldots \mathrm{BCA}_{\downarrow}\mathrm{CABCABC}\ldots$$

This is called an *intrinsic stacking fault*. It is formed, for instance, by the precipitation of a layer of vacancies, but it is also generated in the wake of a glissile Shockley partial dislocation with Burgers vector $\frac{1}{6}[11\bar{2}]$ on a (111) glide plane.

If a single atomic layer is inserted, when for instance a layer of interstitials precipitates, the resulting sequence is

$$\ldots \mathrm{ABC}^{\downarrow}\mathrm{BABCABC}\ldots$$

This is called an *extrinsic stacking fault*. The displacement vector is now $\boldsymbol{R}_0 = -\frac{1}{3}[111]$ (i.e., the opposite of the previous one). A single glide dislocation cannot generate such a fault. In both faults two triplets in the hexagonal configuration occur, but in a different configuration.

To make a detailed interpretation of partial dislocation–fault configurations in face-centered-cubic metals, it is important to distinguish between intrinsic and extrinsic stacking faults. It has been shown that this information can be obtained from the nature of the edge fringes in stacking fault images [82–84]. In particular [85], it has been demonstrated how this information can be obtained from a single dark field image made in a well-defined reflection.

One can distinguish three classes of reflections in the face-centered-cubic structure, depending on whether or not $h+k+l =$ Three-fold, Three-fold + 1, or Three-fold − 1. The reflections for which $h+k+l =$ Three-fold lead to $\alpha = k \cdot 2\pi$ and, therefore, do not give rise to a fringe pattern. Reflections such as {200}, $\{2\bar{2}0\}$, and {440}, for which $h+k+l =$ Three-fold − 1, will be called type A, whereas reflections such as $\{1\bar{1}1\}$, {220}, and {400}, for which $h+k+l =$ Three-fold + 1, will be called type-B reflections.

The edge fringes in bright and dark field images for all possible combinations of the type of active reflection (A or B), the sense of inclination of the fault planes, and the nature of the fault (E or I) are represented schematically in Table 2 where the diffraction vector is assumed to point to the right.

Table 2. Determination of the type of stacking fault in the face-centered-cubic structure.

Bright field				Dark field			
A		B		A		B	
D	D	B	B	B	D	D	B
B	B	D	D	B	D	D	B
B	B	D	D	D	B	B	D
D	D	B	B	D	B	B	D

$\vec{g}$

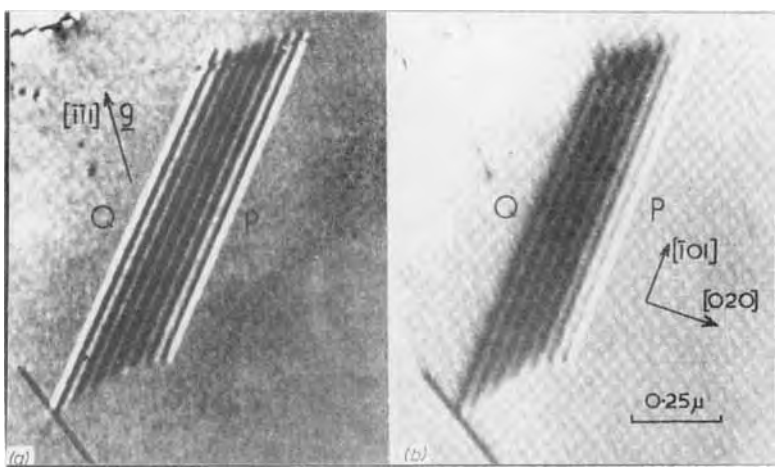


Figure 112. Bright (a) and dark field (b) image pair of a stacking fault in a small stacking fault energy alloy. From Table 2 it can be concluded that the fault is intrinsic.

Note that for a given type of fault and a given type of vector the nature of the edge fringes in the dark field image is independent of the sense of inclination of the fault plane. A simple rule can thus be formulated: if in the dark field image the $\boldsymbol{g}$ vector, its origin being put in the center of the fringe pattern, points towards a bright fringe and the operating reflection $\boldsymbol{g}$ is of type A, then the fault is intrinsic. If one of the parameters (the nature of the edge fringes or the class of the operating reflection) changes, the conclusion also changes. The nature of the edge fringes in the bright

field image also allows one to determine the sense of inclination of the fault plane. In applying the present method one must be aware of the fact that the nature of the edge fringes is only well defined in sufficiently thick foils, where anomalous absorption is important.

Figure 112 shows a bright and dark field image pair which allows us to conclude that the fault being imaged is an intrinsic fault.

Acknowledgements

Thanks are due to the numerous colleagues for providing photographs for illustrations.

Reference is made in the captions to the original publications or in the case of unpublished photographs to the original author. In particular Prof. G. Van Tendeloo provided several unpublished photographs.

1.1.1.21 References

[1] N. F. Mott, H. S. W. Massey, *The Theory of Atomic Collisions*, Clarendon, Oxford, **1949**.
[2] J. A. Ibers, B. K. Vainshtein, in *International Crystallographic Tables*, Volume III, Tables 3.3.3 A(1) and A(2), Kynock Press, Birmingham, **1962**.
[3] W. L. Bragg, *Nature* **1929,** *124*, 125.
[4] P. P. Ewald, *Ann. Phys.* **1917,** *54*, 519.
[5] S. Takagi, *Acta Crystallogr.* **1962,** *15*, 1311.
[6] S. Kiguchi, *Jpn J. Phys.* **1928,** *5*, 23.
[7] C. G. Darwin, *Phil. Mag.* **1914,** *27*, 315, 675.
[8] G. Thomas, in *Diffraction and Imaging Techniques in Material Science* (Eds.: S. Amelinckx, R. Gevers, J. Van Landuyt), North-Holland, Amsterdam, **1978,** p. 217.
[9] R. Gevers, in *Diffraction and Imaging Techniques in Material Science* (Eds.: S. Amelinckx, R. Gevers, J. Van Landuyt), North-Holland, Amsterdam, **1978,** p. 9.
[10] P. B. Hirsch, A. Howie, M. J. Whelan, *Phil. Mag.* **1962,** *7*, 2095.
[11] P. B. Hirsch, A. Howie, M. J. Whelan, *Phil. Trans. R. Soc., Ser. A* **1960,** *252*, 499.
[12] R. Gevers, *Phil Mag.* **1962,** *7*, 59, 651, 769, 1681.
[13] A. Howie, M. J. Whelan, in *Proc. European Reg. Conf. on Electron Microscopy, Delft*, Vol. 1, **1960,** p. 194.
[14] A. Howie, M. J. Whelan, *Proc. R. Soc., Ser. A* **1961,** *263*, 206, 217.
[15] A. Amelinckx, D. Van Dyck, *Electron Diffraction Techniques*, Vol. 2, IUCr Monographs on Crystallography 4, Oxford Science Publications, Oxford, **1993,** p, 000.
[16] C. G. Darwin, *Phil. Mag.* **1914,** *27*, 315, 675.
[17] S. Amelinckx, in *Ullmann's Encyclopedia of Industrial Chemistry*, Vol. B6, VCH, Weinheim, **1994,** p. 229.
[18] S. Amelinckx, in *Solid State Physics* (Eds.: F. Seitz, D. Turnbull), Academic Press, London, **1964,** p. 405.
[19] S. Amelinckx, in *Solid State Physics* (Eds.: F. Seitz, D. Turnbull), Academic Press, London, **1964,** p. 409.
[20] J. W. Menter, *Proc. R. Soc., Ser. A* **1956,** *236*, 119.
[21] S. Amelinckx, *J. Electron. Microscop. Technol.* **1986,** *3*, 131.
[22] H. Yoshioka, *J. Phys. Soc. Jpn* **1957,** *12*, 628.
[23] G. Borrmann, *Phys. Z.* **1941,** *42*, 157; **1950,** *127*, 297.
[24] S. Amelinckx, J. Van Landuyt, in *Diffraction and Imaging Techniques in Material Science* (Eds.: S. Amelinckx, R. Gevers, J. Van Landuyt), North-Holland, Amsterdam, **1978,** p. 107.
[25] R. Gevers, J. Van Landuyt, S. Amelinckx, *Phys. Stat. Solidi* **1965,** *11*, 689.
[26] G. Van Tendeloo, S. Amelinckx, *Phys. Stat. Solidi A* **1974,** *22*, 621.
[27] G. Van Tendeloo, S. Amelinckx, *Acta Crystallogr., Sect. A* **1974,** *30*, 431.
[28] J. Van Landuyt, G. Van Tendeloo, S. Amelinckx, M. B. Walker, *Phys. Rev. B* **1985,** *31*, 2986.
[29] G. Dolino, P. Bachheimer, B. Berge, C. M. Zeyen, G. Van Tendeloo, J. Van Landuyt, S. Amelinckx, *J. Phys.* **1984,** *45*, 901.
[30] G. Van Tendeloo, J. Van Landuyt, S. Amelinckx, *Phys. Stat. Solidi A* **1976,** *33*, 723.
[31] M. Meulemans, P. Delavignette, F. Garcia-Gonzales, S. Amelinckx, *Mat. Res. Bull.* **1970,** *5*, 1025.
[32] C. Boulesteix, J. Van Landuyt, S. Amelinckx, *Phys. Stat. Solidi A* **1976,** *33*, 595.
[33] C. Manolikas, J. Van Landuyt, S. Amelinckx, *Phys. Stat. Solidi A* **1979,** *53*, 327.
[34] R. Serneels, M. Snijkers, P. Delavignette, R. Gevers, S. Amelinckx, *Phys. Stat. Solidi B* **1973,** *58*, 277.
[35] J. D. Eshelby, A. N. Stroh, *Phil. Mag.* **1951,** *42*, 1401.

[36] M. Snijkers, R. Serneels, P. Delavignette, R. Gevers, S. Amelinckx, *Cryst. Latt. Def.* **1972,** *3*, 99.
[37] M. Mannami, *J. Phys. Soc. Jpn* **1962,** *17*, 1160.
[38] M. Mannami, *Acta Crystallogr.* **1960,** *13*, 363.
[39] P. Delavignette, R. Trivedi, R. Gevers, S. Amelinckx, *Phys. Stat. Sol.* **1966,** *17*, 221.
[40] M. F. Ashby, L. M. Brown, *Phil. Mag.* **1963,** *8*, 1083, 1649.
[41] M. Wilkens, in *Diffraction and Imaging Techniques in Material Science* (Eds.: S. Amelinckx, R. Gevers, J. Van Landuyt), North-Holland, Amsterdam, **1978,** p. 185.
[42] U. Essmann, M. Wilkins, *Phys. Stat. Solidi* **1964,** *4*, K53.
[43] D. J. H. Cockayne, M. J. Jenkins, I. L. E. Ray, *Phil. Mag.* **1971,** *24*, 1383.
[44] D. J. H. Cockayne, I. L. E. Ray, M. J. Whelan, *Phil. Mag.* **1969,** *20*, 1265.
[45] R. De Ridder, S. Amelinckx, *Phys. Stat. Solidi B* **1971,** *43*, 541.
[46] R. Gevers, *Phys. Stat. Solidi* **1963,** *3*, 415.
[47] P. Humble, in *Diffraction and Imaging Techniques in Material Science* (Eds.: S. Amelinckx, R. Gevers, J. Van Landuyt), North-Holland, Amsterdam, **1978,** p. 315.
[48] A. K. Head, P. Humble, L. M. Clarebrough, A. T. Morton, C. T. Forwood, 'Computed Electron Micrographs and Defect Identification', in *Defects in Crystalline Solids*, Vol. 7 (Eds.: S. Amelinckx, R. Gevers, J. Nihoul).
[49] P. Delavignette, R. Trevedi, R. Gevers, S. Amelinckx, *Phys. Stat. Sol.* **1966,** *17*, 221.
[50] R. Gevers, *Phys. Stat. Solidi* **1963,** *3*, 2289.
[51] D. W. Pashley, J. W. Menter, G. A. Bassett, *Nature* **1957,** *179*, 752.
[52] G. A. Bassett, J. W. Menter, D. W. Pashley, *Proc. R. Soc., Ser. A* **1958,** *246*, 345.
[53] F. R. N. Nabarro (Ed.), *Dislocation in Solids*, North-Holland, Amsterdam, **1979.**
[54] R. Siems, P. Delavignette, S. Amelinckx, *Phys. Stat. Solidi* **1962,** *2*, 421, 636.
[55] R. Siems, P. Delavignette, S. Amelinckx, *Z. Phys.* **1961,** *165*, 502.
[56] S. Amelinckx, P. Delavignette, *J. Appl. Phys.* **1962,** *33*, 1458.
[57] S. Amelinckx, in *Dislocation in Solids* (Ed.: F. R. N. Nabarro), North-Holland, Amsterdam, **1979,** Vol. 2, p. 345.
[58] J. Van Landuyt, G. Remaut, S. Amelinckx, *Phys. Stat. Solidi* **1970,** *41*, 271.
[59] E. Ruedl, P. Delavignette, S. Amelinckx, *J. Nucl. Mater.* **1962,** *6*, 46.
[60] A. Berghezan, A. Fourdeux, S. Amelinckx, *Acta Metall.* **1960,** *9*, 464.
[61] S. Amelinckx, in *Chemistry and Physics of Carbon*, Vol. I (Ed.: P. L. Walker), Marcel Dekker, New York, **1965,** p. 1.
[62] P. B. Hirsch, J. Silcox, R. Smallmann, K. Westmacott, *Phil. Mag.* **1958,** *3*, 897.
[63] P. B. Hirsch, R. W. Horne, M. J. Whelan, *Phil. Mag.* **1956,** *1*, 667.
[64] P. B. Hirsch, J. Silcox, in *Growth and Perfection of Crystals* (Eds.: R. H. Doremus et al.), Wiley, New York, **1958,** p. 262.
[65] H. Hashimoto, A. Howie, M. J. Whelan, *Proc. R. Soc., Ser. A* **1962,** *269*, 80.
[66] A. Art, R. Gevers, S. Amelinckx, *Phys. Stat. Solidi* **1963,** *3*, 967.
[67] M. J. Whelan, P. B. Hirsch, *Phil. Mag.* **1957,** *2*, 1121; ibid., 1303.
[68] G. Van Tendeloo, S. Amelinckx, *Acta Crystallogr., Sect. A* **1974,** *30*, 431. R. Portier, D. Gracia, *Journal de Physique* **1982,** *43*, C4-17.
[69] C. Manolikas, J. Van Landuyt, S. Amelinckx, *Phys. Stat. Solidi A* **1979,** *53*, 327.
[70] C. Manolikas, J. Van Landuyt, S. Amelinckx, *Phys. Stat. Solidi A* **1980,** *60*, 607.
[71] C. Manolikas, S. Amelinckx, *Phys. Stat. Solidi* **1980,** *61*, 179.
[72] H. W. Zandbergen, G. Van Tendeloo, T. Okabe, S. Amelinckx, *Phys. Stat. Solidi A* **1987,** *103*, 45.
[73] P. Delavignette, S. Amelinckx, *Appl. Phys. Lett.* **1963,** *2*, 236.
[74] S. Amelinckx, *Chimica Scripta* **1978, 1979,** *14*, 197.
[75] G. Van Tendeloo, S. Amelinckx, *Phys. Stat. Solidi A* **1974,** *22*, 621.
[76] D. Broddin, G. Van Tendeloo, J. Van Landuyt, S. Amelinckx, *Phil. Mag.* **1989,** *59*, 47.
[77] D. Broddin, G. Van Tendeloo, S. Amelinckx, *J. Phys. Condens. Matter* **1990,** *2*, 3459.
[78] G. Van Tendeloo, S. Ghose, S. Amelinckx, *Phys. Chem. Minerals* **1989,** *16*, 311.
[79] J. Vanhellemont, S. Amelinckx, *J. Appl. Phys. B* **1987,** *61*, 2176.
[80] J. Mahy, J. Van Landuyt, S. Amelinckx, Y. Uchida, K. D. Bronsema, S. Smaalen, *Phys. Rev. Lett.* **1985,** *55*, 1181.
[81] J. Mahy, J. Van Landuyt, S. Amelinckx, K. D. Bronsema, S. Smaalen, *J. Phys. C: Solid State Physics* **1986,** *C19*, 5049.
[82] H. Hashimoto, A. Howie, M. J. Whelan, *Proc. R. Soc.* **1962,** *A267*, 80.
[83] H. Hashimoto, M. J. Whelan, *J. Phys. Soc. Jpn* **1963,** *18*, 1706.
[84] R. Gevers, A. Art, S. Amelinckx, *Phys. Stat. Solidi* **1963,** *3*, 1563.
[85] A. Art, R. Gevers, S. Amelinckx, *Phys. Stat. Solidi* **1963,** *3*, 967.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

1.1.2 High-Resolution Electron Microscopy

1.1.2.1 Introduction

Among all the techniques used to obtain structural information of materials, high-resolution electron microscopy (HREM) has the great advantage that it yields information about the bulk structure, projected along the direction of electron incidence at a resolution comparable to the interatomic distances. This enables the study of complicated and artificial structures, crystal defects, precipitates and so forth, down to the atomic level. It is our belief that in the future to come, when materials research will gradually evolve into materials design and microstructures into nanostructures, the disposal of a high-resolution electron microscope yielding quantitative structural information at the subangstrom level will be indispensable.

By combining the information from different projections one can in principle obtain three-dimensional structural information. For instance, in the case of silicium, a resolution of 0.1 nm allows observation of the crystal along seven different zone axes.

Recent technological improvements allow one to obtain a resolution of about 0.1 nm, so that it will become possible to 'see' the individual atomic columns of which matter is constituted. However, the potential power of the technique is still severely limited by the problem of quantitative interpretation of the images. For instance, the use of computer simulation images requires much a priori knowledge, which makes HREM very dependent on other techniques. The situation can be compared with the early days of X-ray diffraction when the power of the technique became obvious but the applications were limited by lack of interpretation. Recent developments make it possible to retrieve the object structure directly from the electron micrographs. In this way HREM becomes much more powerful and independent. We will discuss future prospects in this direction and we will also show that the ultimate resolution is finally limited by the object itself.

1.1.2.2 Principles of Image Formation

Basic Principles

Let us first consider, as an illustrative example, the simplest imaging device: the camera obscura. This is a black box with a pinhole (Fig. 1). The object, represented by the function $f(x)$, is projected through the

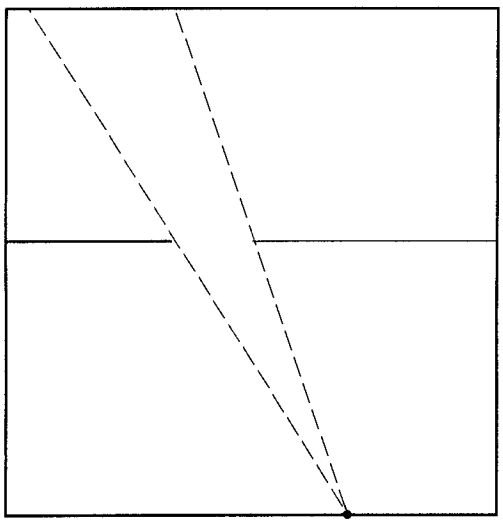


Figure 1. The simplest imaging device: the projection box.

pinhole (aperture) to the image (for simplicity we take the function and the camera to be one-dimensional). A point x in the image gets contributions from an area of the object as seen through the window. This can be mathematically expressed as

$$f_{\text{im}}(x) = \int a(x' - x) f(x')\, \mathrm{d}x' \qquad (1)$$

where $a(x)$ is the aperture function, which is equal to one in the aperture and zero elsewhere. Equation (1) is the definition of a convolution product

$$f_{\text{im}} = a(x) * f(x) \qquad (2)$$

If we take the Fourier transform of Eq. (2), we obtain

$$\tilde{f}_{\text{im}}(g) = \tilde{a}(g)\tilde{f}(g) \qquad (3)$$

i.e. the Fourier transform of a convolution product is the product of the Fourier transforms. g is the spatial frequency. This is illustrated on the right-hand side of Fig. 2. $\tilde{a}(g)$ is usually called the (modulation) transfer function or MFT of the imaging device. It is shown schematically in Fig. 3.

Every imaging device can be characterized by its transfer function (band filter), which describes the magnitude with which a spatial frequency g is transferred through the device. The noise, N, is also indicated.

Resolution

Usually, the resolution of the instrument ρ is defined from the cut-off $1/\rho$ between the signal and noise beyond which no spatial information is transferred. This is the type of resolution in the sense as defined by Rayleigh. The Fourier transform of the transfer function to real space is usually called the impulse response function (IRF). It is the generalization of the aperture function of the camera obscura. It is a sharply peaked function which represents the image of a point object. The width of the IRF is also related to the Rayleigh resolution. The sharper the IRF, the better the resolution. This is demonstrated in Fig. 4. If the transfer function is known, the original image can be restored up to the resolution ρ by dividing by $\tilde{a}(g)$. This is called image restoration or deblurring. If an imaging device consists of a series of subdevices, the total transfer function is the product of the transfer functions of the subdevices.

1.1.2.3 The Electron Microscope [1]

Transfer Function

The image formation in an electron microscope is a coherent process, so that the object function as well as the transfer

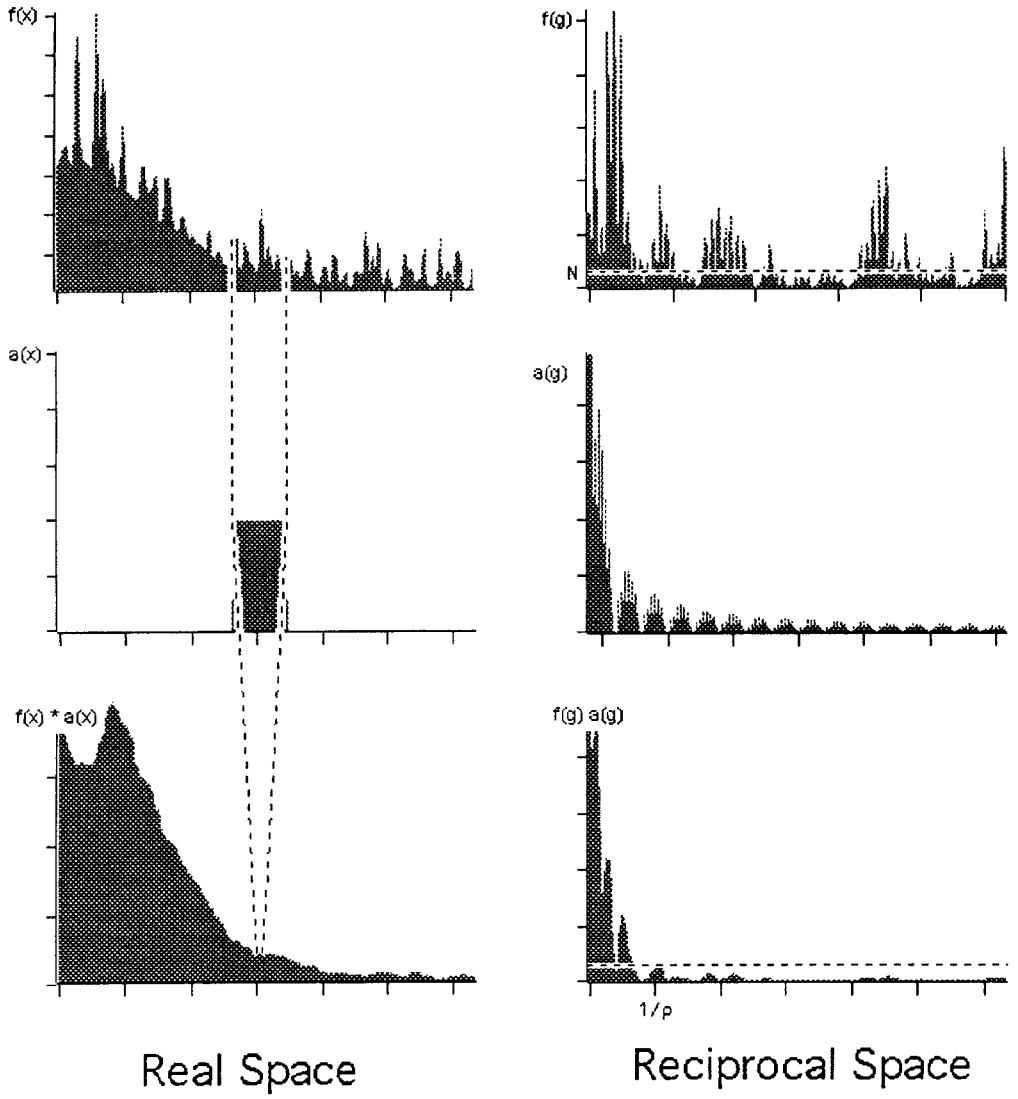


Figure 2. Mathematical formulation of the image formation in a projection box in real space (left) and reciprocal space (right).

functions are complex functions with an amplitude and a phase component. The transfer function now has a particular physical meaning.

The wavefunction $\psi(\boldsymbol{R})$ at the exit face of the object can be considered as a planar source of spherical waves (Huyghens principle) ($\boldsymbol{R}$ is taken in the plane of the exit face). It thus acts as a diffraction grating. According to Fraunhofer's diffraction theory the complex amplitude of the diffracted wave in the direction given by the reciprocal vector $\boldsymbol{g}$ (or spatial frequency) is given by the Fourier transform

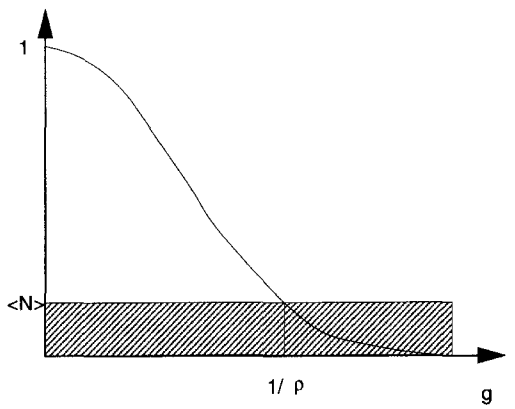


Figure 3. Transfer function.

of the object function, i.e.

$$\psi(\boldsymbol{g}) = F_{\boldsymbol{g}}\psi(\boldsymbol{R}) \tag{4}$$

Now the objective lens is placed behind the object. Hence it focuses each parallel diffracted beam into another point of the focal plane, whose position is given by the reciprocal vector $\boldsymbol{g}$ characterizing the diffracted beam. Thus the back focal plane of the objective lens contains the Fourier transform of the object. In an electron microscope it is possible to change the lens settings so as to image the focal plane directly. In this way one can see the diffraction pattern, given by $|\psi(\boldsymbol{g})|^2$. If the object is periodic, such as a crystal, the diffraction pattern will consist of sharp spots. A continuous object will give rise to a continuous diffraction pattern. In the second stage of the imaging process, the back focal plane acts, in its turn, as a set of Huyghens sources of spherical waves which interfere, through a system of lenses, in the image plane. This stage in the imaging process is described by an inverse Fourier transform which reconstructs the object function $\psi(\boldsymbol{R})$ (usually enlarged) in the image plane (Fig. 5). The intensity in the image plane is then given by $|\psi(\boldsymbol{R})^2|$.

In practice, by inserting an aperture in the focal plane of the objective lens, it is possible to obtain an image in which only selected beams contribute (see Sec. 1.1.2.5). By selecting only one beam, all interference disappears and the image shows only contrast proportional to the local intensity of that particular diffraction beam. This mode of imaging is called diffraction contrast (see Sect. 1.1.1 of this Chapter).

The electron microscope however is not a perfect imaging device. On passing through the objective lens, each electron beam $\boldsymbol{g}$ (i.e. the spatial frequency g) undergoes a phase shift and an amplitude reduction (damping). Hence the transfer function takes the form

$$T(\boldsymbol{g}) = A(\boldsymbol{g})\exp[-\mathrm{i}\chi(\boldsymbol{g})]D(\alpha, \Delta, \boldsymbol{g}) \tag{5}$$

$\chi(g)$ is the phase shift. $A(g)$ describes the effect of the beam selecting aperture and the damping. For a derivation of the exact expression, see Appendix A. The wavefunction at the image plane is now given by

$$\Phi(\boldsymbol{R}) = F_{\boldsymbol{R}}^{-1}T(\boldsymbol{g})\psi(\boldsymbol{g}) \tag{6}$$

and the image intensity by

$$I(\boldsymbol{R}) = |\phi(\boldsymbol{R})|^2 \tag{7}$$

Equation (5) is called the coherent approximation; it is valid for thin objects. For thicker objects one uses the concept of transmission cross-coefficient (TCC). Here, the Fourier components of the image intensity are given by

$$I(\boldsymbol{g}) = F_{\boldsymbol{g}}(I(\boldsymbol{R})) = \int \psi(\boldsymbol{g}+\boldsymbol{g}')\tau(\boldsymbol{g}+\boldsymbol{g}', \boldsymbol{g}')\psi^*(\boldsymbol{g}')\,\mathrm{d}\boldsymbol{g}' \tag{8}$$

with τ the TCC which describes how the beams $\boldsymbol{g}'$ and $\boldsymbol{g}+\boldsymbol{g}'$ are coupled in the

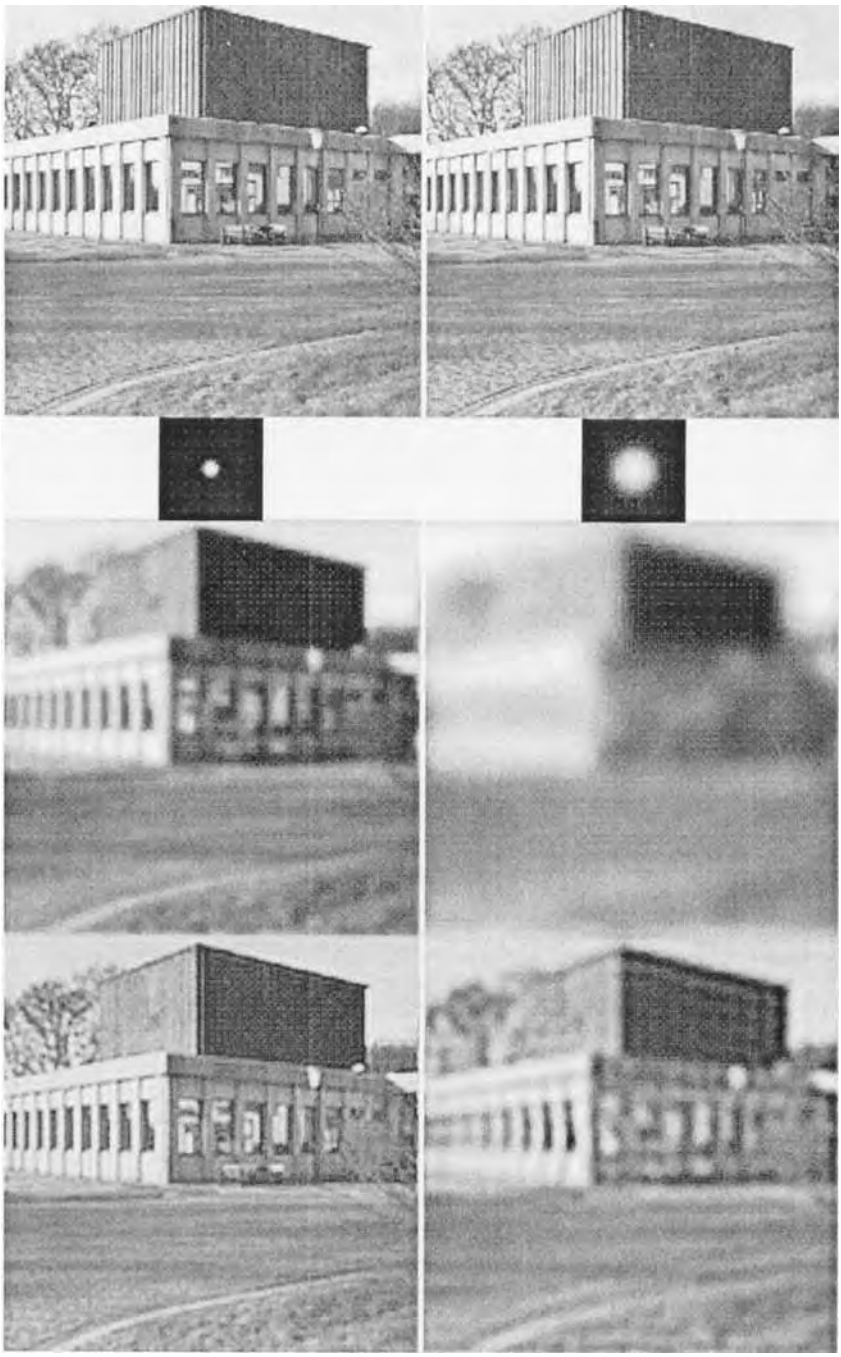


Figure 4. Image formation through a pinhole (impulse response function). From top to bottom: original image (Centre for Electron Microscopy, Antwerp); Gaussian pinholes with different sizes; blurred image, deblurred image.

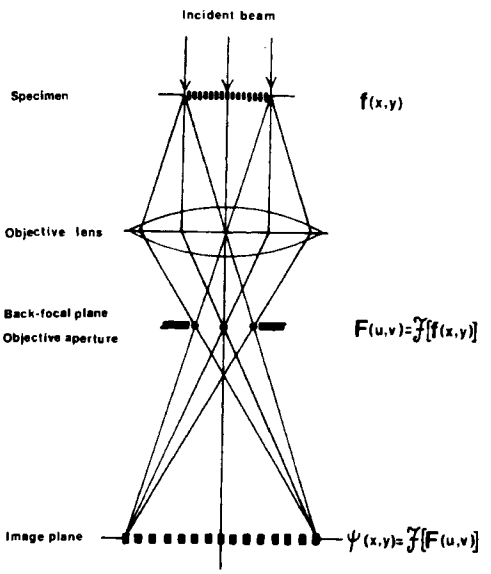


Figure 5. Schematic representation of the image formation by the objective lens in a transmission electron microscope. The corresponding mathematical operations are indicated (see text with $\bar{R} = (x, y)$ and $\bar{G} = (u, v)$).

Fourier component $I(\boldsymbol{g})$. For a derivation of Eq. (8) see Appendix A.

The total image formation process is shown schematically in Fig. 6.

Impulse Response Function

As explained in Sec. 1.1.2.2 the image transfer can also be described in real space as a convolution product

$$\Phi(\boldsymbol{R}) = \psi(\boldsymbol{R}) * t(\boldsymbol{R}) \tag{9}$$

where $\psi(\boldsymbol{R})$ is the object wave in real space and $t(\boldsymbol{R})$ is the Fourier transform of the transfer function. For a hypothetical ideal pointlike object, $\psi(\boldsymbol{R})$ would be a delta function or 'impulse' so that $\Phi(\boldsymbol{R}) = t(\boldsymbol{R})$, that is, the microscope would reveal $t(\boldsymbol{R})$, which is therefore called the impulse response function. This is the equivalent of the window function in Fig. 2. If the transfer function was constant (i.e., perfectly flat) over the whole spatial frequency range, the impulse response would be a delta function so that $\Phi(\boldsymbol{R}) = \psi(\boldsymbol{R})$, that is, the wavefunction in the image plane represents exactly the wavefunction of the object. In a sense the microscope is perfect. However, in practice the transfer function cannot be made arbitrarily flat as is shown in Fig. 7. The impulse response function is still peaked as shown in Fig. 8.

Hence, as follows from Eq. (9), the object wavefunction $\Phi(\boldsymbol{R})$ is then smeared out (blurred) over the width of the peak. This width can then be considered as a measure for the resolution in the sense as originally defined by Rayleigh. The width of this peak is the inverse of the width of the constant plateau of the transfer

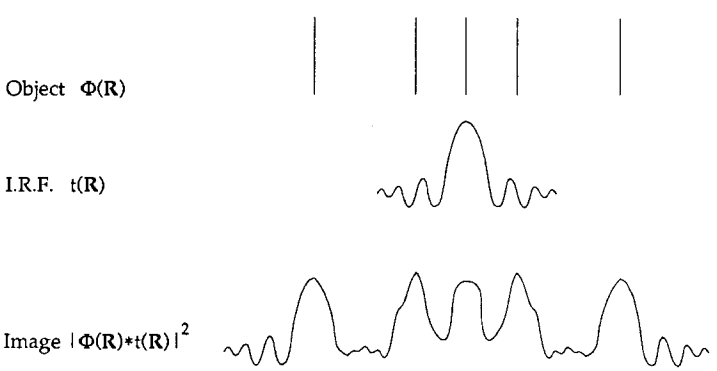


Figure 6. Schematic representation of the whole image formation process in the electron microscope.

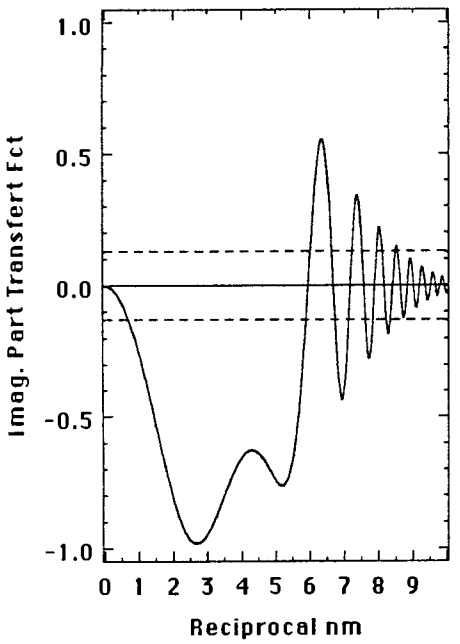


Figure 7. Transfer function (imaginary part) of a 300 keV electron microscope at optimum defocus.

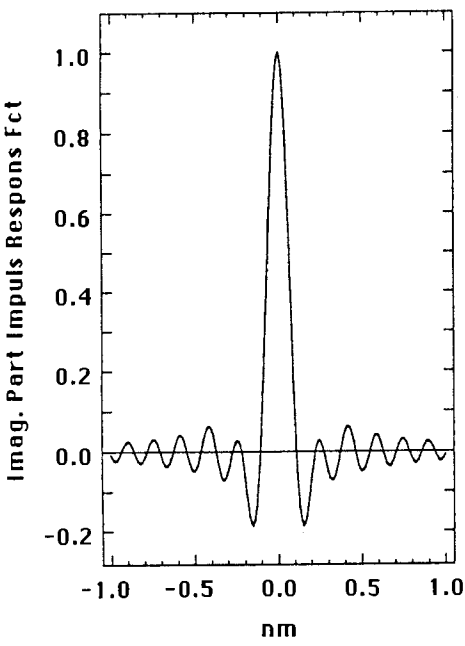


Figure 8. Impulse response function (imaginary part) corresponding to the transfer function of Fig. 6.

function in Fig. 7. From another point of view one can argue that if all the spatial frequencies have the same phase shift, the information is transferred forward and keeps a point-to-point relation to the object.

However, the information beyond this plateau is still contributing to the image but with a wrong phase. It is scattered outside the peak of the impulse response function and thus redistributed over a larger area in the image plane.

Imaging at Optimum Focus: Phase Contrast Microscopy

In an ideal microscope, the image wavefunction would exactly represent the object wavefunction, and the image intensity for a pure phase object function would be

$$|\Phi(\boldsymbol{R})|^2 = |\psi(\boldsymbol{R})|^2 = |\exp[\mathrm{i}\varphi(\boldsymbol{R})]|^2 = 1 \tag{10}$$

that is, the image would show no contrast. This can be compared with imaging a glass plate with variable thickness in an ideal light microscope.

Assuming a weak phase object (WPO) one has $\varphi(\boldsymbol{R}) \ll 1$ so that

$$\psi(\boldsymbol{R}) \approx 1 + \mathrm{i}\varphi(\boldsymbol{R}) \tag{11}$$

The constant term 1 contributes to the central beam (zeroth Fourier component) whereas the term $\mathrm{i}\varphi$ mainly contributes to the diffracted beams. If the phases of the diffracted beams can be shifted over $\pi/2$ with respect to the central beam, the amplitudes of the diffracted beams are multiplied by $\exp(\mathrm{i}\pi/2) = \mathrm{i}$. Hence the image term $\mathrm{i}\varphi(\boldsymbol{R})$ becomes $-\varphi(\boldsymbol{R})$. It is as if the

object function had the form

$$\Phi(\boldsymbol{R}) = 1 - \varphi(\boldsymbol{R}) \approx \exp[-\varphi(\boldsymbol{R})]$$

that is, the phase object now acts as an amplitude object. The image intensity is then

$$|\Phi(\boldsymbol{R})|^2 \approx 1 - 2\varphi(\boldsymbol{R}) \tag{12}$$

which is a direct representation of the phase of the object. In light microscopy, this has been achieved by F. Zernike using a quarter-wavelength plate.

In electron microscopy, phase contrast imaging can be achieved by making the transfer function as constant as possible. From Eq. (37) it is clear that oscillations occur due to spherical aberration and defocus. However, the effect of spherical aberration which, in a sense, makes the objective lens too strong for the most inclined beams, can be compensated somewhat by slightly underfocusing the lens. The optimum defocus value (also called Scherzer* defocus) for which the plateau width is maximal, is given by

$$\epsilon_s = -1.2(\lambda C_s)^{1/2} = -1.2 \text{ Sch} \tag{13}$$

with $1\,\text{Sch} = (\lambda C_s)^{1/2}$ the Scherzer unit, with λ the electron wavelength and C_s the spherical aberration. Typical values are $\lambda = 2\,\text{pm}$ (300 keV) and $C_s = 1\,\text{mm}$, yielding $\epsilon_s = -54\,\text{mm}$.

The transfer function for this situation is depicted in Fig. 7. The phase shift $\chi(g)$ is nearly equal to $-\pi/2$ for a large range of spatial coordinates g. The Scherzer plateau extends nearly to the first zero at a spatial frequency

$$g \approx 1.5 C_s^{-1/4} \lambda^{-3/4} \tag{14}$$

* Named after Otto Scherzer who was the first to describe the image formation in the electron microscope in this way [1].

In this mode one can image directly the phase of the object. Now a thin material object acts as a phase object (see Appendix B) in which the phase is proportional to the electrostatic potential of the atoms projected along the viewing direction. Hence, if the object was very thin, optimum focus imaging would directly reveal atoms as dark areas, and empty spaces as light. The details are explained in Sec. 1.1.2.6. However, this argument only holds for distances which are well above the point resolution of the microscope. Furthermore, the thickness up to which an object can be considered as WPO is unrealistically small (e.g., 1 nm) and is rarely met in practice.

Resolution of the Electron Microscope

In principle the characteristics of the electron microscope can be completely defined by its transfer function, that is, by the instrumental aberrations. However, a clear definition of resolution is not easily given for an electron microscope. For instance, for thick specimens, there is not necessarily a one-to-one correspondence between the projected structure of the object and the wavefunction at the exit face of the object, so that the image does not show a simple relationship.

If one wants to determine a 'resolution' number, this can only be meaningful for thin objects. Furthermore one has to distinguish between *point resolution* (or *structural resolution*), as the finest detail that can be interpreted in terms of the structure, and the *information limit*, which is the

finest detail that can be resolved by the instrument, irrespective of a possible interpretation.

As explained in Sec. 1.1.2.6, the electron microscope in the phase contrast mode at optimum focus directly reveals the projected potential, that is, the structure, of the object, provided the object is very thin. All spatial frequencies g with a nearly constant phase shift are transferred forward from object to image. Hence the point resolution can be obtained from the first zero of the transfer function (14) as

$$\rho_s = \frac{1}{g} \approx 0.65 C_s^{1/4} \lambda^{3/4} = 0.65\,\mathrm{Gl} \qquad (15)$$

$\mathrm{Gl} = C_s^{1/4}\lambda^{3/4}$ is called the Glaser unit. The point resolution is also equal to the 'width' of the impulse response function. The information beyond ρ_s is transferred with a non-constant phase and, as a consequence, is redistributed over a larger image area.

The *information limit* can be defined as the finest detail that can be resolved by the instrument. It corresponds to the maximal diffracted beam angle that is still transmitted with appreciable intensity, that is, the transfer function of the microscope is a spatial band filter which cuts all information beyond the information limit. For a thin specimen, this limit is mainly determined by the envelope of chromatic aberration (temporal incoherence) and beam convergence (spatial incoherence) (see Appendix A). In principle, beam convergence can be reduced using a smaller illuminating aperture and a larger exposure time. If chromatic aberration is predominant, the damping envelope function is given by Eq. (40), from which the information limit can be estimated as

$$\rho_I = \frac{1}{g} = \left(\frac{\pi\lambda\Delta}{2}\right)^{1/2} \qquad (16)$$

The information limit is usually smaller than the point resolution. Typical values are $\lambda = 2\,\mathrm{pm}$ (300 keV), $C_s = 1\,\mathrm{mm}$, $\Delta = 5\,\mathrm{nm}$, $\rho_s = 0.2\,\mathrm{nm}$, and $\rho_I = 0.13\,\mathrm{nm}$. The point resolution can be improved by reducing C_s and reducing λ (i.e., increasing the voltage). The information limit can be improved by improving the coherence and reducing λ.

1.1.2.4 Resolution Limits [2]

The electron microscope can be considered as an information channel that transfers information from the object to the observer. The channel can be considered as consisting of four subchannels, each of which has a transfer function that limits the resolution.

These are schematically represented in Fig. 9. We will discuss each of the four different subchannels in more detail.

The Atom

The electron interacts with the electrostatic potential of the atom. A typical electrostatic potential of an atom has a Gaussian-like shape. The average width is of the order of 0.05–0.1 nm. The Fourier transform of the potential is also Gaussian shaped, and is called the scattering factor. The atom itself can be considered as the ultimate probe with which the object can be sampled. So the final resolution in the

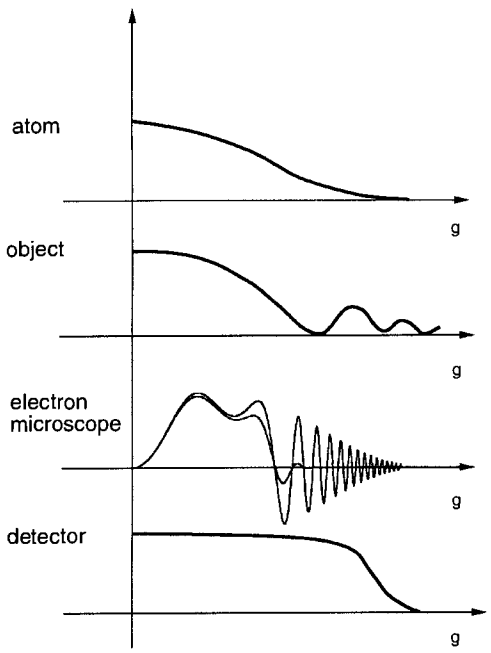


Figure 9. Transfer functions of the subchannels in the electron microscope.

sense of Rayleigh will always be of the order of 0.1 nm. Furthermore, the thermal atom motion causes the scattering factor to decrease by the so-called Debye–Waller factor, which even reduces the resolution.

The Object

In practice the imaging in the electron microscope is two-dimensional whereas the object itself is three-dimensional. In a sense the three-dimensional structure is forced into a two-dimensional imaging process. This also has an effect on the resolution. The conservation of energy requires that the wave vector of the electrons should be on the Ewald sphere. Hence the scattered intensity is reduced with increasing distance from the origin.

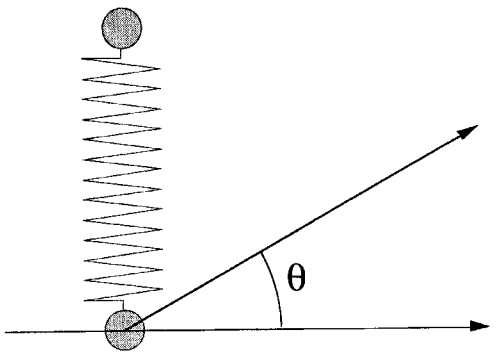


Figure 10. Schematic representation of electron scattering at a configuration of two atoms connected by a spring.

The ultimate resolution limit imposed by the object can be estimated as follows. By studying an object one is in fact interested in knowing the configuration of atoms. However, in order to get information with sufficient detail, the incident electron has to approach the atoms closely so that it can transmit energy, change the configuration, and destroy the coherence. This poses ultimate intrinsic limits to the resolution. We will demonstrate it in a very simplified way.

Consider two neighboring atoms connected by a spring (Fig. 10). The energy transferred by the incident electron is given by the formula

$$\frac{\Delta E}{E} = \frac{4m}{M} \sin^2\left(\frac{\theta}{2}\right) \tag{17}$$

with M the mass of the target atom. The diffraction angle is related to the resolution by the Bragg formula

$$\rho = \frac{\lambda}{\sin\theta} \tag{18}$$

The energy of the spring considered as a harmonic oscillator is given by

$$\Delta E = \frac{h}{2\pi}\sqrt{\frac{k}{M}} \tag{19}$$

with

$$k = \frac{dK}{dr} \tag{20}$$

K is the Coulomb force and

$$E = \frac{h^2 k^2}{2m} \tag{21}$$

Finally one obtains, using Eqs. (17)–(21), for the ultimate resolution

$$\rho = \frac{\epsilon R_B}{Z^{1/4}} \tag{22}$$

with R_B the Bohr radius, Z the atomic number, and ϵ a configuration constant approximately equal to 1.5. Hence, the electrons that carry out information beyond ρ will transfer energy to the system and are lost for coherent imaging. The ultimate resolution for coherent imaging is thus of the order of 0.1 nm for light atoms and 0.05 nm for heavy atoms. Note also that this limit is independent of the type of the imaging particles.

The Electron Microscope

The ultimate resolution is determined by the subchannel with the worst resolution. Thus far, the weakest part has been the electron microscope itself.

As shown in Sec. 1.1.2.3 the point resolution ρ_S of the electron microscope can be improved by reducing the spherical aberration constant C_s and/or by increasing the voltage. However, since C_s depends mainly on the pole piece dimension and the magnetic materials used, not much improvement can be expected. Hence, at present, all high-resolution electron microscopes yield comparable values for C_s for comparable designs. Furthermore, the effect of C_s on the resolution is rather limited. In the far future, a major improvement can be expected by using superconducting lenses or microlenses.

Another way of increasing the resolution is by correcting the third-order spherical aberration by means of a system of quadrupole, hexapole, and/or octupole lenses [3].

Increasing the voltage is another way of increasing the resolution. However, apart from the increased cost, increasing the voltage might increase the displacive radiation damage of the object, especially for light atoms.

A promising way of increasing the resolution is by restoring the information that is present between ρ_S and ρ_I and that is still present in the image, albeit with the wrong phase. For this purpose, holography combined with image processing will be indispensable. In that case, the resolution will be determined by ρ_I. ρ_I can be improved drastically by using a field emission gun (FEG) which reduces the spatial as well as the temporal incoherence. Taking all these considerations into account, an ultimate resolution of the electron microscope of 0.1 nm is within reach even at intermediate voltages.

Since resolution is a trade-off between signal and noise, some improvement can still be expected by reducing the noise. Specimen noise (inelastic scattering) can be reduced by energy filtering and the recording noise can be improved by using slowscan CCD cameras. However, if we assume that the total transfer function is Gaussian, an improvement in the signal-to-noise ratio from 20 to 100 only results in an improvement of the resolution with 25%. Ultimately, one cannot avoid the Poisson noise due to the counting statistics (quantum noise or shot noise) which is

limited by the radiation dose. One cannot arbitrarily increase the dose without increasing the recording time and/or damaging the object. Hence, it can be expected that the ultimate resolution attainable with this technique will not exceed 0.05 nm. At that resolution level it will not be the microscope that will limit resolution but rather the object itself. It is worth noticing that when testing the resolution capabilities of an electron microscope close to 0.1 nm, it is hard to find objects with which this resolution can be demonstrated. Conversely, only suitable objects can benefit from a further resolution improvement. Hence we believe that more effort has to be done to optimize the specimen preparation technique.

The Recording Device

If one wants to record information up to the information limit, special requirements have to be met by the recording device. First in real space one needs at least three pixels per resolution unit. In reciprocal space the finest oscillations of the transfer functions (Fig. 7) have to be sampled sufficiently, which means about eight pixels per oscillation. The oscillations can be reduced over the whole spatial frequency range by using an optimum focus called the Lichte focus (see Sec. 1.9 of this Chapter). Totally this gives a minimum number of pixels given by the formula

$$N > 30\left(\frac{\rho_S}{\rho_I}\right)^4 \qquad (23)$$

One now has $N > 500$ for $\rho_S = 0.2$ nm and $\rho_I = 0.1$ nm. This requires a very good recording system. If one uses off-axis holography (see Sec. 1.9 of this Chapter) then this number has to be multiplied by three. The newest generation of slow-scan CCD cameras with fiber-coupled scintillators might be a solution to this problem. Furthermore, when cooled, these cameras are able to detect nearly all single electrons.

'Seeing' Atoms: A New Situation

It is clear that with the recent technological improvements the resolution approaches the physical limits, which is comparable to the size of the atoms (strictly speaking one does not 'see' individual atoms but rather individual columns in projection).

If atoms, the building blocks of matter, can be seen individually, one can introduce a new concept of resolution, based on the information theory of Shannon. In 1965, Gabor argued that one could not use information theory for microscopy since this theory was devised to establish communication between a sender and a receiver which communicated with a common vocabulary of messages. However, in the microscope one communicates between an unknown object and an observer. Since the object is unknown one cannot define a set of messages. However, when it is possible to see the atoms, they can be considered as messages in the original sense of Shannon and hence we can use communication theory. Consider for example the simplified case depicted in Fig. 11. d is a one-dimensional periodicity containing atoms. For each atom the position has to be determined. The right-hand side shows the transfer function. It is sampled in distances $1/d$. The total number of degrees of freedom is equal to the total number of

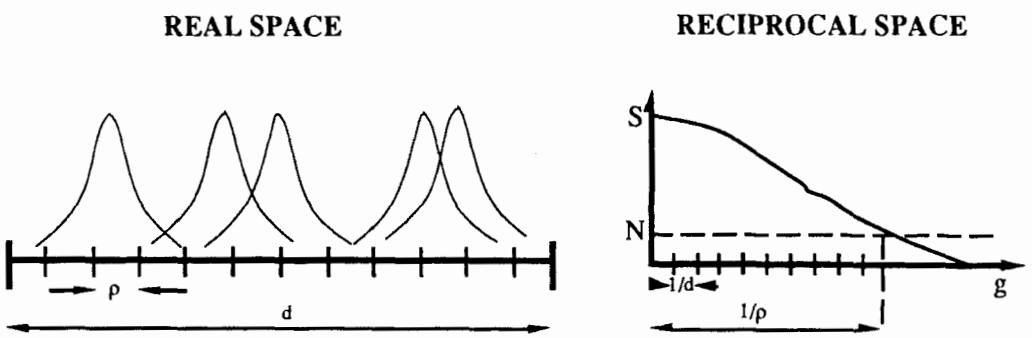


Figure 11. Schematic representation of the information content.

spatial frequencies in the transfer function, given by

$$n = \frac{1/\rho}{1/d} = \frac{d}{\rho} \tag{24}$$

from which

$$\frac{n}{d} = \frac{1}{\rho}$$

Hence each unit of length ρ carries one degree of freedom. The average information of each degree expressed in bits can be calculated as follows: calling S the signal and N the noise (Fig. 11), if we consider the noise as the smallest piece of information. The number of bits needed to describe the signal is then

$$\log_2\left(\frac{N+S}{N}\right) = \log_2\left(1+\frac{S}{N}\right)$$

Strictly speaking this means that a microscope has an information capacity of one number per resolution unit. Each number has on the average $\log_2(1 + S/N)$ bits of information (precision). In two dimensions one can use the same arguments (Fig. 12). For a resolution ρ the number of degrees of freedom per unit area is equal to about three per ρ^2, which in fact means for an electron microscope with a resolution of 0.1 nm that one can determine approximately the coordinates of one atom per 0.01 nm^2. If one investigates a crystal viewed along a simple zone axis and the number of columns in projection is less than one per ρ^2, the complete structure in projection can be solved. If, however, the number of atom columns in projection exceeds the capacity of the microscope, as when one investigates a complex zone

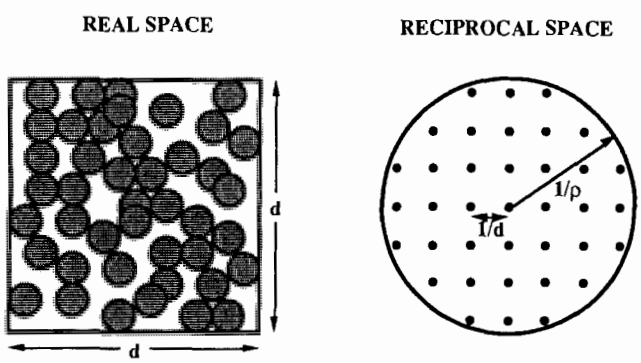


Figure 12. Information content in two-dimensional images.

axis or when the object is disordered, then the capacity of the microscope is exceeded and the information channel is blocked. In such a situation a structure cannot be solved completely and only limited information can be obtained from the images. It is thus more appropriate to regard the resolution capacity of the electron microscope as the number of data points per unit area. The situation is comparable to the case of X-ray diffraction. If the number of beams exceeds the number of atom coordinates to be determined (usually by a factor of two to three) then the structure can in principle be solved completely.

1.1.2.5 High-Resolution Electron Microscopy in Practice

The Electron Microscope

The resolution capability of an electron microscope is related to the size of the objective pole piece gap and thus also to the tilting possibilities. In cases where tilting is not necessary, as often is the case for the study of biological specimens, one can use a high-resolution stage with a very small bore, but without tilting possibilities. Generally, for materials research, a goniometer stage is indispensable. This can either be a double tilt or a rotation-tilt stage, although the former is easier to use. Moreover, one has to make a compromise between the maximal tilt angle and the resolution. The higher the resolution, the smaller the tilt angle and the more patience is required from the operator to find a suitable orientation. However, at high voltages ($\geqslant 300$ keV) it is possible to combine high resolution with large tilt possibilities. One can usually choose between a top entry and a side entry goniometer stage. The top entry stage normally yields a higher resolution and better mechanical stability (less thermal specimen drift). The side entry stage has much better possibilities for tilting, heating, strain experiments, and so on, usually at the expense of resolution. However, recently designed side entry stages seem to compete with top entry stages with respect to resolution and stability.

For high-resolution work many advantages are offered if the electron microscope is equipped with an image-intensifying camera mounted below the photographic plate holder. This facilitates the observation of the high-resolution images and the adjustment of the microscope. A suitable rule of thumb for making good use of the camera is that the resolution of the camera, divided by the primary magnification of the electron microscope, should be much smaller than the resolution of the electron microscope. For instance for 300 keV electrons the resolution of the best CCD camera is of the order of 50 μm (mainly determined by the resolution of the transmission fluorescent screen). Therefore, a primary magnification of 10^6 is required in order that $50\,\mu m \times 10^{-6} = 0.05\,nm$ should be much less than the resolution of the electron microscope.

Before starting high-resolution work, it is necessary to determine the most important optical parameters of the instrument for later use in image simulation and reconstruction. For very high resolution (below 0.2 nm) the standard correction procedure for the aberrations is not sufficient, and methods have been

developed for automatic alignment. Furthermore, the spherical aberration constant depends on the position of the specimen in the objective lens. For this purpose the microscope should be equipped with a CCD camera and on-line image-processing capabilities. A series of images is taken from a thin amorphous film at different focus values. Each of the images is then digitized and numerically Fourier transformed so as to obtain the so-called diffractogram. Originally, diffractograms were obtained by diffraction with a laser beam in an optical bench, but due to the disposal of good CCD cameras and fast computers this method is not much in use anymore. The diffractogram represents the contrast transfer function corresponding with the particular focus. Accurate values for the spherical aberration constant C_s and for the defocus values can be derived by curve fitting. In this way the defocus steps can be calibrated. The point resolution can be measured from the first zero of the transfer function and from the damping function. The instrumental resolution, the chromatic aberration, and the beam convergence can also be determined.

A review of aberrations and their determination is given by Saxton [4]. Recently, it has been discovered [5] that third-order astigmatism, which is usually believed to be unimportant, may become very important for very high resolution. The effect of this astigmatism is not visible in the diffractogram. Recently it has been shown how it can be measured and corrected [4]. It is also interesting to note that when holographic reconstruction methods are used (see Sec. 1.1.2.6) the microscope has not to be tuned perfectly since the residual aberrations can be corrected for afterwards. One of the major problems encountered in obtaining high-resolution images is the mechanical stability. During the exposure (of the order of seconds) the drift of the specimen has to be much less than the resolution of the instrument. This can be overcome somewhat by decreasing the primary magnification so that the intensity is increased and the recording time reduced. The resolution and sensitivity of the photographic plate or the camera then are the limiting factors.

The Specimen

Specimens for HREM are prepared using the same techniques as for usual transmission electron microscopy, that is, ion beam milling, chemical and electrolytical thinning, cleavage, crushing, and so forth. The only requirements are that the specimen should be sufficiently thin, that is, less than about 10 nm. Furthermore, the specimen should be sufficiently clean and free of contamination. For instance, when a specimen such as an alloy is thinned using ion milling, it can be recommended to finish with a chemical polishing or plasma cleaning, in order to remove the amorphous layer.

For brittle specimens, the crushing method is usually applied. Here the specimens are ground in an agate mortar after which the powder is suspended in methanol and some drops of this suspension are deposited on a thin perforated supporting film.

In cases where cleavage along a required plane does not occur easily, it might be worthwhile to grind the specimens at liquid nitrogen temperature where the material is more brittle and the cleavage more isotropic. In cases where the required cleavage is almost impossible, for example

when hair-grown crystals have to be cleaved perpendicular to the hair filaments, one can try to embed the crystals in an epoxy resin and to cut them by microtomy.

For details of specimen preparation we refer the reader to Robards and Wilson [6].

Crystalline specimens with a unit cell with two large and one small lattice parameter are most ideal for HREM. In that case the reciprocal lattice consists of dense planes (Laue zones) which are largely separated. Such crystals can be oriented with their short axis parallel to the incident beam so that the nearly flat Ewald sphere touches the Laue zone through the origin (Fig. 13). Hence a large number of diffracted beams are excited simultaneously and maximal information is present in the image. This is shown in Fig. 14. In this situation, the electrons propagate parallel to a zone axis, that is, parallel to the atom rows. Only in this way can a possible interpretation of the images in terms of the projected structure be meaningful. The same argument holds also for crystals with defects.

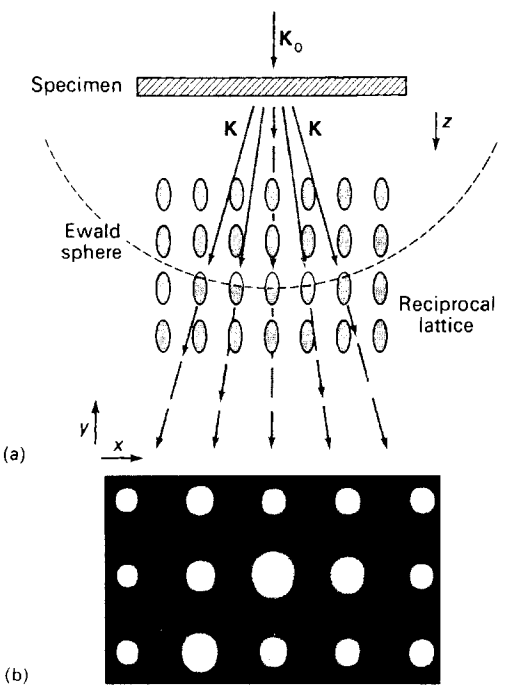


Figure 13. Formation of the diffraction pattern. The simultaneously excited electron beams can be used for image formation.

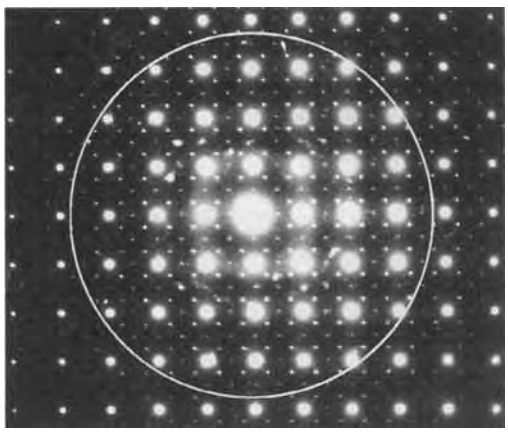


Figure 14. Typical diffraction pattern as used for structure imaging. The aperture selecting the contributing beams is also indicated.

As an example, Fig. 15 shows a schematized model of a dislocation viewed along the atomic columns.

It is also possible, using an aperture placed in the focal plane of the objective lens, to select a particular set of beams so that the images contain specific

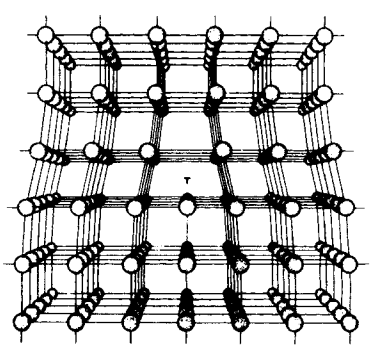


Figure 15. Structure model of a crystal containing a dislocation viewed along the atomic columns.

information. If the central beam is not included, these images are called dark field images. Examples are given in Sec. 1.1.2.7.

When the periodicity along the incident beam direction is not small, the Laue zones are not distant from each other so that more Laue zones are excited at the same time. In this case, the diffraction pattern also shows the intersection of the Ewald sphere with the higher-order Laue zones (HOLZs), which consist of concentric circles. In that case the three lattice parameters can be deduced from one diffraction pattern.

When dealing with a crushed specimen suspended on holey support film, one has to search for thin, wedge-shaped flakes, partly covering a hole of the film. Then the specimen has to be oriented with the required zone axis parallel to the incident beam. However, when the tilt possibilities are not large (e.g., 10°) the majority of the specimens cannot be properly oriented, except when the cleavage is preferentially perpendicular to the zone axis. It can then be tried to search for a suitable crystal particle in diffraction mode. A crystal with a nearly good orientation can easily be recognized since in that case the intersection of the Ewald sphere with the zeroth Laue zone is an arc of a circle through the origin, which, with some experience, can easily be recognized. The smaller the radius of curvature of the arc, the closer to the exact zone orientation. After finding a suitably thin part with the proper orientation, one has to adjust the focus. When the specimen is very thin, the zero focus corresponds to minimal contrast (see Sec. 1.1.2.6). By underfocusing, a bright fringe appears at the border of the specimen. Maximal contrast appears close to the optimum defocus. On going through focus, the contrast reverses.

In practice, since the focus is not exactly known, especially in the case of thicker specimens, one has to take a series of images at gradually different focus settings, recorded approximately around the optimum defocus. This is called a through-focus series (Fig. 16). When a CCD camera is available, the number of recorded through-focus images can usually be reduced. When dealing with a specimen that is unstable in the electron beam, the specimen can be completely destroyed within a period of seconds. Here a minimal exposure technique has to be used. In some cases, only one picture can be taken. Here also the availability of an image intensifier allows use of a much smaller beam intensity and an increase in the time of observation.

A parameter that is difficult to determine is the specimen thickness. In the case of a crystalline specimen with a small unit cell, an estimate of the thickness in thicker areas can sometimes be obtained from thickness contours in a two-beam case. Tilt experiments can also provide some information. However, in general the thickness of the foil in thin areas cannot be measured accurately.

1.1.2.6 Interpretation of the Images

Intuitively Interpretable Images

Images of Thin Objects

(i) *Optimum focus*

When the phase change in the phase object is very small (weak phase object) we can use expression (56) derived in Appendix B,

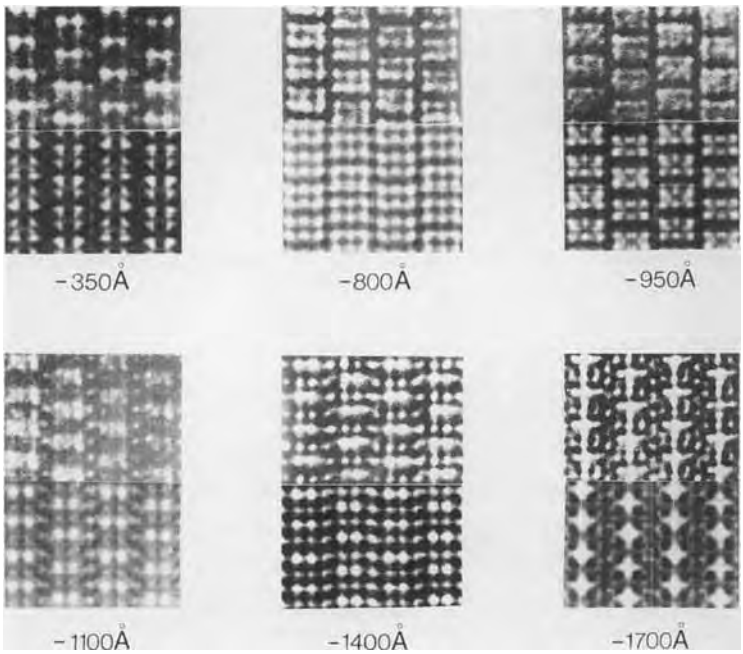


Figure 16. Comparison of experimental images (top row) (courtesy of S. Iijima) and computer-simulated images (bottom row) for $Ti_2Nb_{10}O_{29}$ as a function of defocus.

which can be expanded up to the first power as

$$\psi(\boldsymbol{R}) \simeq 1 + \mathrm{i}\sigma V_\mathrm{p}(\boldsymbol{R}) - \mu(\boldsymbol{R}) \tag{25}$$

$V_\mathrm{p}(\boldsymbol{R})$ is the projected potential of the object, $\mu(\boldsymbol{R})$ is the absorption function. The Fourier transform, yielding the amplitude in the back focal plane now becomes

$$\psi(\boldsymbol{g}) = \delta(\boldsymbol{g}) + \mathrm{i}\sigma V_\mathrm{p}(\boldsymbol{g}) - M(\boldsymbol{g}) \tag{26}$$

with the Dirac function $\delta(\boldsymbol{g})$ representing the transmitted beam.

From Equation (5) the image amplitude (without aperture) is now

$$\begin{aligned}\phi(\boldsymbol{R}) &= \mathscr{F}_{\boldsymbol{R}}\psi(\boldsymbol{g})\,\mathrm{e}^{-\mathrm{i}\chi(\boldsymbol{g})} \\ &= \mathscr{F}_{\boldsymbol{R}}[\delta(\boldsymbol{g}) + \sigma V_\mathrm{p}(\boldsymbol{g})\sin\chi(\boldsymbol{g}) \\ &\quad - M(\boldsymbol{g})\cos\chi(\boldsymbol{g}) \\ &\quad + \mathrm{i}\sigma V_\mathrm{p}(\boldsymbol{g})\cos\chi(\boldsymbol{g}) \\ &\quad + \mathrm{i}M(\boldsymbol{g})\sin\chi(\boldsymbol{g})]\end{aligned} \tag{27}$$

At the optimum defocus the transfer function shows a nearly flat region for which $\sin\chi(\boldsymbol{g}) \approx -1$ and $\cos\chi(\boldsymbol{g}) \approx 0$ for all contributing beams.

Now Eq. (27) becomes

$$\begin{aligned}\phi(\boldsymbol{R}) &\approx \mathscr{F}_{\boldsymbol{R}}[\delta(\boldsymbol{g}) - \sigma V_\mathrm{p}(\boldsymbol{g}) - \mathrm{i}M(\boldsymbol{g})] \\ &= 1 - \sigma V_\mathrm{p}(\boldsymbol{R}) - \mathrm{i}\mu(\boldsymbol{R})\end{aligned}$$

and the image intensity to the first order is

$$I(\boldsymbol{R}) \approx 1 - 2\sigma V_\mathrm{p}(\boldsymbol{R}) \tag{28}$$

At the optimum focus, the electron microscope acts as a phase contrast microscope so that the image contrast of a thin object is proportional to its electrostatic potential $V_\mathrm{p}(\boldsymbol{R})$ projected along the direction of incidence. This theory can be generalized for larger phase changes [7].

(ii) *Small Defocus*

In the absence of apertures and spherical aberration, the image amplitude is

given by

$$\phi(\boldsymbol{R}) = \mathcal{F}_{\boldsymbol{R}} \exp[-i\pi\epsilon\lambda(\boldsymbol{g}^2)]\psi(\boldsymbol{g}) \tag{29}$$

which can be expanded for small defocus values as

$$\begin{aligned}\phi(\boldsymbol{R}) &= \mathcal{F}_{\boldsymbol{R}}[1 - i\pi\epsilon\lambda(\boldsymbol{g}^2)]\psi(\boldsymbol{g}) \\ &= \psi(\boldsymbol{R}) - i\pi\epsilon\lambda\, \mathcal{F}_{\boldsymbol{R}}(\boldsymbol{g}^2)\psi(\boldsymbol{g})\end{aligned}$$

or explicitly

$$\begin{aligned}\phi(\boldsymbol{R}) = \psi(\boldsymbol{R}) &- i\pi\epsilon\lambda \\ &\times \int \exp[2\pi i(\boldsymbol{g}\cdot\boldsymbol{R})](\boldsymbol{g}^2)\psi(\boldsymbol{g})\, d\boldsymbol{g}\end{aligned}$$

This expression can be transformed elegantly as

$$\begin{aligned}\phi(\boldsymbol{R}) &= \psi(\boldsymbol{R}) + \frac{i\pi\epsilon\lambda}{4\pi^2}(\nabla^2) \\ &\quad \times \int \exp[2\pi i(\boldsymbol{g}\cdot\boldsymbol{R})]\psi(\boldsymbol{g})\, d\boldsymbol{g} \\ &= \left(1 + \frac{i\epsilon\lambda}{4\pi}(\nabla^2)\right)\psi(\boldsymbol{R})\end{aligned}$$

which for a phase object, where $\psi(\boldsymbol{R}) = \exp[i\sigma V(\boldsymbol{R})]$, becomes

$$\begin{aligned}\phi(\boldsymbol{R}) = \Bigg(1 &- \frac{\epsilon\lambda\sigma\Delta V_p(x,y)}{4\pi} \\ &- \frac{i\epsilon\lambda\sigma^2}{4\pi}(\nabla V(\boldsymbol{R}))^2\Bigg) \\ &\times \exp[i\sigma V_p(\boldsymbol{R})]\end{aligned}$$

So that the image density $|\phi(\boldsymbol{R})|^2$ is given to the first order by

$$I(\boldsymbol{R}) = 1 - \frac{\epsilon\lambda\alpha\Delta V_p(\boldsymbol{R})}{2\pi} \tag{30}$$

with Δ the Laplacean operator. Since the projected $V_p(\boldsymbol{R})$ is of electrostatic origin, it obeys the Poisson equation,

$$\Delta V_p(\boldsymbol{R}) = -4\pi\rho_p(\boldsymbol{R})$$

so that finally

$$I(\boldsymbol{R}) = 1 + 2\epsilon\lambda\sigma\rho_p(\boldsymbol{R}) \tag{31}$$

The image contrast is thus proportional to the projected charge density; it disappears in the Gaussian image plane ($\epsilon = 0$) and reverses with the sign of the defocus. This effect can be used to find approximately the zero focus. Expression (31) is the projected charge density (PCD) approximation [8]. Its validity is restricted to resolutions for which the spherical aberration is not too important. It is interesting to note that exactly at zero focus the image of the phase object shows no contrast. The only contrast observed stems from absorption, by which electrons are removed from the image formation. The use of a small objective aperture which eliminates electrons from the imaging process also gives rise to a kind of absorption contrast. This is the case for biological samples, 'colored' with heavy atoms.

For phase objects at small underfocus, neglecting the spherical aberration, the image contrast shows the projected charge density. For WPOs at optimum defocus, including the spherical aberration, the image contrast represents the projected potential. Although both results are in fact approximations, it can be concluded that between zero focus and optimum defocus, the details of the high-resolution images of very thin objects which are larger than the point resolution can be interpreted intuitively in terms of the projected specimen structure. For instance, groups of atoms are imaged with black contrast, and holes are imaged with white contrast. Examples are given in Sec. 1.1.2.7. However, these cases are rarely met in practice.

Interpretation Using Channeling

The wavefunction in the image plane can be written as the convolution product of the wavefunction at the exit face of the crystal with the impulse response function $t(\boldsymbol{R})$ of the electron microscope, Eq. (9). When we now use the expression derived for the wavefunction using the channeling theory (Eq. (89)), we obtain

$$\psi(\boldsymbol{R}) = 1 + \sum_i \left[\exp\left(\frac{-\mathrm{i}E_i}{E}\,\frac{z}{\lambda}\right) - 1\right] \times C_i\phi_i(\boldsymbol{R} - \boldsymbol{R}_i) * t(\boldsymbol{R}) \qquad (32)$$

The summation runs over the columns i of the object, parallel to the electron beam with position $\boldsymbol{R}_i$. Each column contributes with a localized wavefunction ϕ_i multiplied with a factor which varies with depth. The periodicity is inversely proportional to E_i which is related to the mass density of the column.

If the microscope is operated close to optimum focus and in axial mode, the impulse response function is sharply peaked (see Fig. 8). If also the localized wavefunction ϕ_i is highly peaked and if the distance between the columns is larger than the width of the impulse response function $t(\boldsymbol{R})$, the overlap between convolution products $\phi_i * t(\boldsymbol{R})$ of adjacent sites can be assumed to be small so that the image intensity is

$$I(\boldsymbol{R}) = |\psi(\boldsymbol{R})|^2 \approx \sum_i 4C_i^2 \sin^2\left(\frac{E_i z}{2E\lambda}\right) \times |\phi_i(\boldsymbol{R} - \boldsymbol{R}_i) * t(\boldsymbol{R})|^2 \qquad (33)$$

Each column is thus imaged separately. The contrast of a particular column varies periodically with thickness. The periodicity can be different for different types of columns. It is interesting to note that the functions ϕ_i as well as $t(\boldsymbol{R})$ are symmetrical around the origin, provided the objective aperture is centered around the optical axis. Hence, the image of a column is rotationally symmetric around the position $\boldsymbol{R}_i$ of the columns. The intensity at $\boldsymbol{R}_i$ is a maximum or a minimum. The positions of the columns can thus be determined from the positions of the intensity extrema.

When the resolution of the microscope is insufficient to discriminate the individual columns, or the focus is not close to optimum, the overlap between the convolution products of adjacent columns cannot be avoided and the interpretation of the contrast is not straightforward. In that case image simulation or object wave reconstruction is required.

Building Block Structures

It often happens that a family of crystal structures exists, all members of which consist of a stacking of simple building blocks but with a different stacking sequence. This is for instance the case in mixed-layer compounds, polytypes, periodical twins, and substitutional columnar binary alloys. In a broader sense also periodic interfaces such as antiphase boundaries and crystallographic shear planes can be considered as building block systems.

In order to characterize the stacking sequence of unknown members of the family, one can search for a member with a known stacking sequence. Often, this will be a simple member, the structure of which has been determined by X-ray diffraction.

From this system, high-resolution micrographs are taken. The image characteristics of each type of building block in these images are then called the imaging code: since the high-energy electrons are scattered forward, it can be expected that the image of a building block will not be severely affected by its surroundings, provided the crystal is not too thick and the focus is close to the optimum focus. Now the building blocks of the unknown members are imaged with the same code, provided the experimental conditions (thickness, focus) are the same, so that the stacking sequence can be read directly from the high-resolution micrographs. In general, the technique of the imaging code is applicable up to crystal thicknesses of a few tens of nanometers, and also to other imaging modes.

Examples are given in Sec. 1.1.2.7.

A particular type of building block structure is the substitutional binary alloys with a column structure. In a substitutional binary alloy, the two types of atoms occupy positions on a regular lattice, usually face-centered cubic (f.c.c.). Since the lattice, as well as the types of the atoms and the average composition, is known, the problem of structure determination is then reduced to a binary problem of determining which atom is located at which lattice site.

Particularly interesting are the alloys in which columns are found parallel to a given direction and which consist of atoms of the same type. Examples are the gold–manganese system and other f.c.c. alloys [9]. If viewed along the column direction, which is usually $[0\,0\,1]_{\text{f.c.c.}}$, the high-resolution images contain sufficient information to determine unambiguously the type and position of the individual columns. Even if the resolution of the microscope is insufficient to resolve the individual lattice positions, which have a separation of about 0.2 nm, it is possible to reveal the minority columns only, which is sufficient to resolve the complete structure. This can be done using dark field superlattice imaging, which is explained in the next section.

Selective Beam Imaging

In some cases it is possible to reveal only specific information by selecting the appropriate beams for contributing to the image function. For this one uses an aperture in the focal plane of the objective lens. An illustrative example is the imaging of binary f.c.c. alloys with a column structure as described above.

Let us consider a binary system consisting of columns of A atoms, which are the majority atoms, and columns of B atoms. We call the fraction of columns respectively m_A and m_B; $m_A + m_B = 1$ with $m_A > m_B$.

Using the channeling theory (see Appendix D), it can then be shown that, when the structure is imaged without the f.c.c., reflections and overlap between images of adjacent columns can be neglected, the image intensity is given by

$$|\psi(\boldsymbol{R})|^2 \approx \sum_i \sigma_i^2 |\Delta\phi(\boldsymbol{R} - \boldsymbol{R}_i) * t(\boldsymbol{R})|^2 \quad (34)$$

where $\Delta\phi = \phi_B - \phi_A$ with ϕ_A (respectively ϕ_B) the wavefunctions of the columns A (respectively B), and σ_i the Flynn occupation parameters defined as $\sigma_i = m_A$ for a B column and $\sigma_i = -m_B$ for an A column. The interpretation of Eq. (34) is now as

follows. The atom columns are imaged as bright peaks, the shape of the peaks being given by $|\Delta\phi(\boldsymbol{R}) * t(\boldsymbol{R})|^2$ and the height of the peaks being proportional to m_A^2 for the B-atom columns and m_B^2 for the A-atom columns.

As a consequence, the minority atoms B are imaged as brighter dots than the majority atoms A, the ratio of brightness being equal to $(m_A/m_B)^2$. When this ratio is large (e.g., >10), the minority atoms will be visible as bright dots on a dark background. An example of this high-resolution dark field imaging applied to the Au_4Mn alloy is given in Sec. 1.1.2.7. A nice example of selective beam imaging is the quantitem technique developed by Ourmadz and co-workers in which they use reflections that are sensitive to the composition of atom columns so as to obtain images from which the composition can be deduced by pattern recognition techniques [10]. A survey of selective imaging methods and applications is given by Amelinckx et al. [11].

Image Simulation

When no obvious imaging code is available, the interpretation of high-resolution images often becomes a precarious problem since, especially at very high resolution, the image contrast can vary drastically with focus. As a typical and historical example, high-resolution images obtained by Iijima for the complex oxide $Ti_2Nb_{10}O_{25}$ with a point resolution of approximately 0.35 nm (100 keV) are shown in Fig. 16 (top row). The structure as reproduced in Fig. 17 consists of a

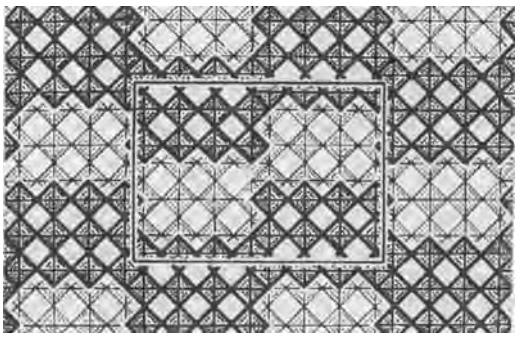


Figure 17. Schematic representation of the unit cell of $Ti_2Nb_{10}O_{29}$ consisting of corner-sharing NbO_6 octahedra with the titanium atoms in tetrahedral sites.

stacking of corner- or face-sharing NbO_6 octahedrons with the titanium atoms in tetrahedral positions. High-resolution images are taken at different focus values, causing the contrast to change drastically. The best resemblance to the structure can be obtained near the optimum Scherzer defocus, which is −90 nm in this particular case. However, the interpretation of such high-resolution images never appears to be trivial. The only way out remains in the comparison of the experimental images with those calculated for various trial structures. During the imaging process, the electrons undergo three distinct interactions, as shown schematically in Fig. 6. Each of these interactions is known and can be calculated by the computer. First, the electron scatters dynamically in the crystal. This interaction can be simulated using the multislice methods explained in Appendix C. However, as an input to the program one has to specify all the object parameters such as the unit cell, position and type of cell atoms, thermal atom factors (Debye–Waller factors), object orientation, and thickness. The result of this calculation yields the wavefunction at

the exit face of the crystal. In a second step, the formation of the image in the electron microscope is simulated using the expressions of Appendix A. Here all the instrumental parameters have to be specified. Finally, the electron intensity in the image plane is calculated by squaring the wavefunction, and is displayed as a halftone image on a high-resolution screen or printer. Different commercial software packages exist for high-resolution image simulations. References are given in Ref. [12].

In practice, the image simulation requires a number of input parameters that are not accurately known such as specimen thickness and focus value. For this reason one often compares whole series of images simulated for various focus values and/or specimen thicknesses which are compared with experimental through-focus and/or through-thickness series, which makes the method more sensitive. As an example, the series of images simulated using the model of Fig. 17 for different focus values are shown in Fig. 16 (bottom row) and reveal a close resemblance to the experimental images. Other examples are given in Sec. 1.1.2.7.

A drawback of the present use of image simulation is that the image comparison is usually done visually and not quantitatively or in a recursive refinement. As a consequence, the technique can only be used if the number of plausible models is very limited. This makes HREM, despite its potential power for structure investigation, very dependent on other techniques. Direct methods, which extract the information from the images in a direct way, are much more promising.

Quantitative Interpretation

In principle one is usually not interested in high-resolution images as such but rather in the structure of object under study. High-resolution images are then to be considered as data planes from which the structural information has to be extracted in a quantitative way. Ideally this should be done as follows. One has a model for the object and for the imaging process, including electron object interaction, microscope transfer, and image detection (see Fig. 6). The model contains parameters that have to be determined by the experiment. The parameters can be estimated from the fit between the theoretical images and the experimental images. The goodness of the fit is evaluated using a matching criterion (fitness function) such as likelihood, least squares, or the R factor (cf. X-ray crystallography). This fitness function can be calculated for each set of parameters. The optimal fit then yields the best estimates for the parameters of the model that can be derived from the experiment. In a sense one is searching for a maximum (or minimum) depending on the criterion) of the fitness function in the parameter space, the dimension of which is equal to the number of parameters. The object model that describes the interaction with the electrons consists of the assembly of the electrostatic potentials of the constituent atoms. Since for each atom type the electrostatic potential is known, the model parameters then reduce to atom numbers and coordinates, thermal atom factors, object thickness, and orientation (if inelastic scattering is neglected).

The imaging process is characterized by a small number of parameters such as

defocus, spherical aberration etc, that are not accurately known.

A major problem is now that the structural information of the object can be strongly delocalized by the image transfer in the electron microscope (see Figs. 6 and 8), so that the influence of the model parameters is completely scrambled in the high-resolution images. Due to this coupling, one has to refine all parameters simultaneously. As a consequence, the dimension of the parameter space is so high that one cannot use advanced optimization techniques such as genetic algorithms, simulating annealing, tabu search, and so forth, without the risk of ending in local maxima. Furthermore, for each new model trial one has to perform a tedious image calculation so that the procedure is very cumbersome. The problem is only manageable if the object is a crystal with a very small unit cell and hence a small number of object parameters [13], or if sufficient prior information is available to reduce the number of parameters drastically.

In X-ray crystallography, this problem can be solved by using direct methods which provide a pathway towards the global maximum. In HREM, this problem can be solved by deblurring the information, so as to unscramble the influence of the different object parameters of the image, and thus reduce the dimension of the parameter space. As described in Sec. 1.1.2.4 this can be achieved by high-voltage microscopy, by correcting the microscopic aberrations, or by holographic techniques.

Holographic techniques have the particular advantage that they first retrieve the whole wavefunction in the image plane, that is, amplitude and phase. In this way, they use all possible information. In the other two methods, one starts from the image intensity only, and the information that is predominantly present in the phase is inevitably missed. Ideally, high-voltage microscopy or aberration correction is combined with holography, which leads to the advantage of holography but with a broader field of view. However, this has not yet been done in practice.

As explained above, the whole purpose is to unscramble the object information in the images, that is, to undo the image formation process, so as to uncouple the object parameters and to reduce the size of the parameter space. In this way it is possible to reach the global maximum (i.e., best fit) which leads to an approximate structure model.

This structure model then provides a starting point for a final refinement by fitting with the original images (i.e., in the high-dimensional parameter space) that is sufficiently close to the global maximum so as to guarantee fast convergence.

It has to be noticed that, in the case of perfect crystals, one can combine the information in the high-resolution images with that of the electron diffraction pattern, which in principle can also be recorded by the CCD camera. Since the diffraction patterns usually yield information up to higher spatial frequencies than the images, one can in this way extend the resolution to beyond 0.1 nm. Recently [14] very accurate structure refinements for unknown structures have been achieved with R factors below 5% (which is comparable to X-ray results). Here one starts with a first estimate of the structure as obtained from exit wave reconstruction (see further) which is then refined iteratively by using the electron diffraction data.

We will now focus attention mainly on the holographic reconstruction methods. Undoing the scrambling from object to

image consists of three stages. First, one has to reconstruct the wavefunction in the image plane (phase retrieval). Then one has to reconstruct the exit wave of the object. Finally, one has to 'invert' the scattering in the object so as to retrieve the object structure.

Phase Retrieval

The phase problem can be solved by holographic methods. Two methods exist for this purpose: off-axis holography and focus variation, which is a kind of in-line holography. In off-axis holography, the beam is split by an electrostatic biprism into a reference beam and a beam that traverses the object. Interference of both beams in the image plane then yields fringes, the positions of which yield the phase information. In order to retrieve this information, a very high-resolution camera (CCD), a powerful image processor, and a field emission source to provide the necessary spatial coherence are needed. Details are given in Sec. 1.9 of this Chapter. In the focus variation method, the focus is used as a controllable parameter so as to yield focus values from which both amplitude and phase information can be extracted [15, 16]. Images are captured at very close focus values so as to collect all the information in the three-dimensional image space. Each image contains linear information and nonlinear information. By Fourier transforming the whole three-dimensional image space, the linear information of all images is superimposed onto a sphere in reciprocal space, which can be considered as an Ewald sphere (Fig. 18). By filtering out this linear

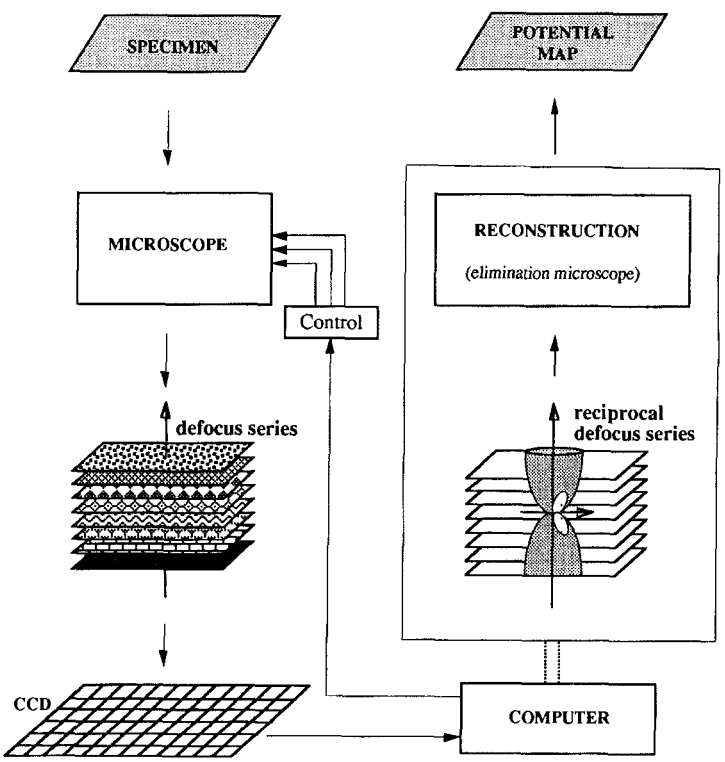


Figure 18. Schematic representation of the phase retrieval procedure. The paraboloid which contains the linear information in reciprocal space is also shown.

information the phase can be retrieved (for details, see Appendix E).

The results indicate that focus variation is more accurate for high spatial frequencies whereas off-axis holography is more accurate for lower spatial frequencies but puts higher demands on the number of pixels in order to detect the high spatial frequencies.

The choice of focal values can also be optimized using a criterion that is currently used for experiment design [17]. It turns out that the choice of equidistant focus values is close to optimal.

Exit Wave Reconstruction

The wavefunction at the exit face of the object can be calculated from the wavefunction in the image plane by applying the inverse phase transfer function of the microscope. This procedure is straightforward, provided the proper parameters describing the transfer function (such as the spherical aberration constant C_s). As is clear from Fig. 7, the retrieval of information up to the information limit requires that the transfer function should be known with high accuracy. This requires an accuracy of less than 0.01 nm for C_s and 5 nm for ϵ. Two remarks have to be made:

(i) In principle the alignment of the microscope has not to be perfect provided the amount of misalignment is known so that it can be corrected for in the reconstruction procedure.

(ii) An accurate measurement of C_s and ϵ can only be performed if sufficient information is known about the object (e.g., a thin amorphous object can be considered as a white noise object) from which the transfer function can be derived from the diffractogram.

We are thus faced with an intrinsic problem. An accurate determination of the instrumental parameters requires knowledge of the object. On the other hand, the most interesting objects under investigation are not fully known. Hence, the fine tuning of the residual aberrations has to be done on the object under study, based on some general assumptions that do not require a knowledge of the specimen structure such as the crystal potential is real, the structure is atomic, and so forth.

For instance, if the object is thin, the phase of the exit wave would show the projected potential which is sharply peaked at the atom columns. If the exit face is reconstructed with a slight residual defocus, these peaks would be blurred. Hence it can be expected that the peakiness of the phase is maximal at the proper defocus. The peakiness can be evaluated by means of an entropy using the Shannon formula. If the object is thicker, it can be expected from the channeling theory (see Eq. (89)) that the amplitude of $\psi - 1$ is peaked, and thus also its entropy. Hence, a weighted entropy criterion may be used for fine tuning the residual defocus. This is shown in Fig. 19. Details are given by Tang et al. [18].

Figure 20 shows the exit wave of an object of $YBa_2Cu_4O_8$ (high T_c superconductor), which was historically the first experimental result obtained with the focus variation method. The microscope used was a Philips CM20 ST equipped with field emission source and $(1024)^2$ slow-scan CCD camera developed in the

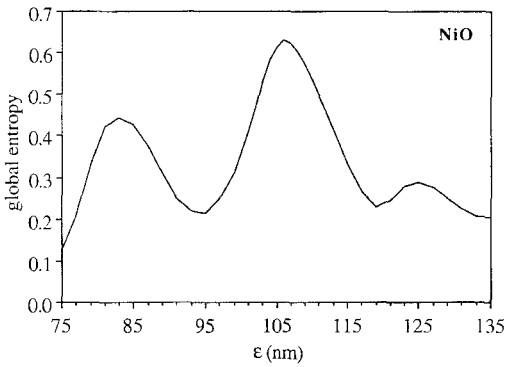


Figure 19. Global exit wave entropy as a function of residual defocus for TiO_2.

framework of a Brite–Euram project. In this case, the object is very thin so that the phase of the wavefunction directly reveals the projected potential of the atom columns. The oxygen columns adjacent to the yttrium columns could just be observed, proving a resolution of 0.13 nm. However, when the object is thicker, the one-to-one correspondence between the wavefunction and the projected structure is not so straightforward due to the dynamic diffraction. This is shown in Fig. 21 for $Ba_2NaNb_5O_{15}$ where the heavy columns (barium and niobium) are revealed in the amplitude and the bright columns (sodium and oxygen) in the phase. In this case it is necessary to invert in a sense the electron scattering in the object so as to retrieve the projected structure.

It should be noted that, once the exit wave is reconstructed, it is in principle possible to recalculate all the images of

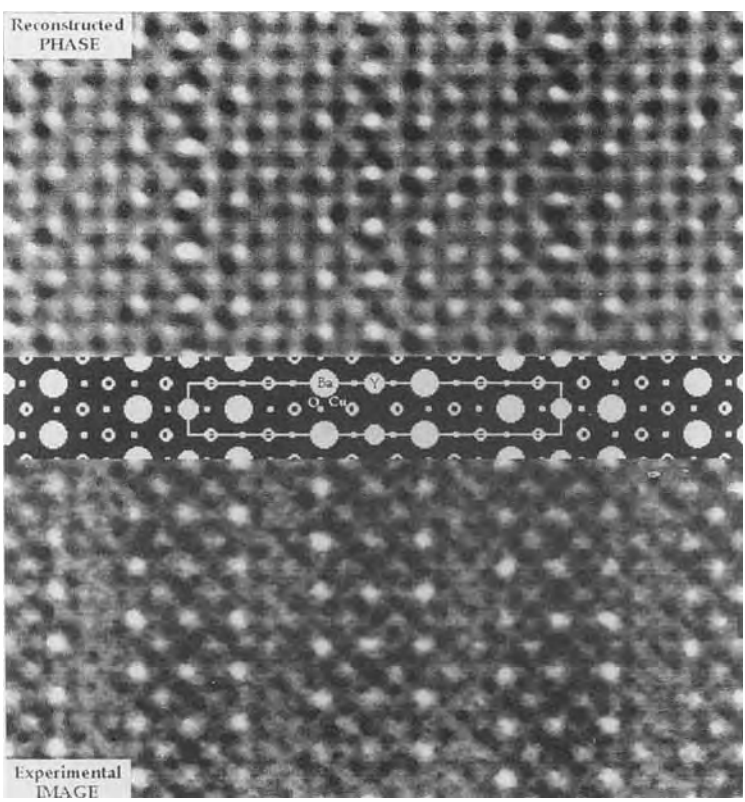


Figure 20. Experimentally reconstructed exit wave for YBa_2CuO_8. Top, reconstructed phase; center, structure model; bottom, experimental image.

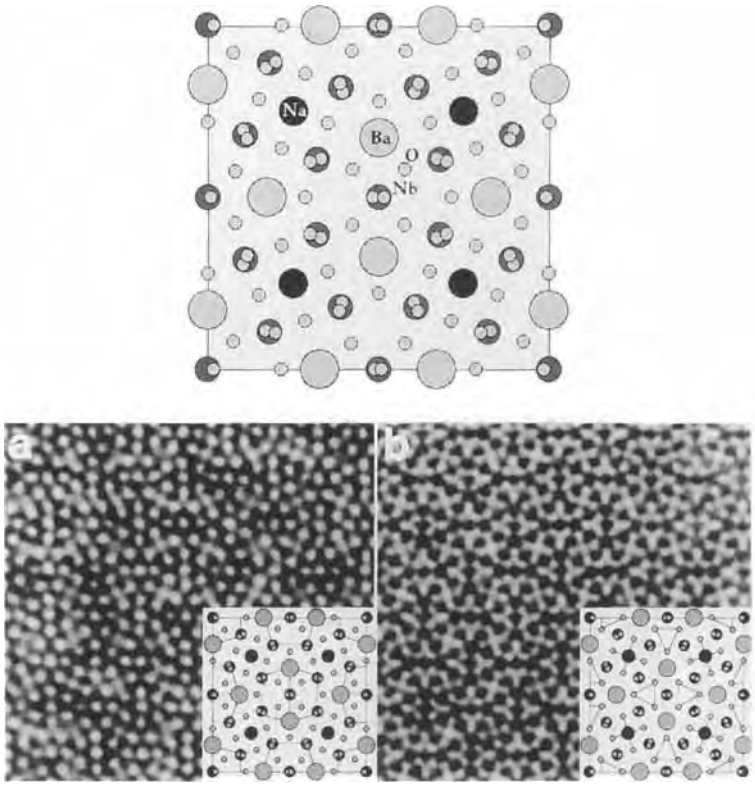


Figure 21. Experimentally reconstructed exit wave for $Ba_2NaNb_5O_{15}$. (a) Amplitude. (b) Phase. The structure model is shown at the top.

the Fourier series which perfectly fit in the experimental images within the noise level. Hence, the reconstructed exit wave contains all experimentally attainable object information. In practice, one thus will not have to store the original images but only the reconstructed wave.

Structure Retrieval

The final step consists of retrieving the projected structure of the object from the wavefunction at the exit face. If the object is thin enough to act as a phase object, the phase is proportional to the electrostatic potential of the structure, projected along the beam direction so that the retrieval is straightforward. If the object is thicker, the problem is much more complicated. In principle one can retrieve the projected structure of the object by an iterative refinement based on fitting the calculated and the experimental exit wave. As explained earlier, this is basically a search procedure in a parameter space. However, since the exit wave is much more locally related to the structure of the object than the original images the dimension of the parameter space is much smaller. Nevertheless it is possible to end in a local maximum [19]. However, an approximate structure can be obtained in a more direct way. If the object is a crystal viewed along a zone axis, the incident beam is parallel to the atom columns. It can be shown that in

such a case the electrons are trapped in the positive electrostatic potential of the atom columns, which then act as channels. This effect is known as electron channeling, and is explained in detail in Appendix D.

If the distance between the columns is not too small, a one-to-one correspondence between the wavefunction at the exit face and the column structure of the crystal is maintained. Within the columns, the electrons oscillate as a function of depth without, however, leaving the column. Hence the classical picture of electrons traversing the crystal as plane-like waves in the direction of the Bragg beams, which historically stems from X-ray diffraction, is in fact misleading. It is important to note that channeling is not a property of a crystal, but occurs even in an isolated column and is not much affected by the neighboring columns, provided the columns do not overlap. Hence the one-to-one relationship is still present in the case of defects such as translation interfaces or dislocations provided they are oriented with the atom columns parallel to the incident beam.

The basic result is that the wavefunction at the exit fact of a column is expressed as Eq. (88):

$$\psi(\boldsymbol{R}, z) = 1 + \left[\exp\left(-i\pi\frac{E}{E_0}kz\right) - 1\right]\phi(\boldsymbol{R}) \tag{35}$$

This result holds for each isolated column. In a sense, the whole wavefunction is uniquely determined by the eigenstate $\phi(\boldsymbol{R})$ of the Hamiltonian of the projected columns and its energy E, which are both functions of the 'density' of the column and the crystal thickness z. It is clear from Eq. (35) that the exit wave is peaked at the centre of the column and varies periodically with depth. The periodicity is inversely related to the 'density' of the column. In this way the exit wave still retains a one-to-one correspondence with the projected structure. Furthermore, it is possible (see Eq. (82)) to parameterize the exit wave in terms of the atomic number Z and the interatomic distance d of the atoms constituting the column. This enables the projected structure of the object to be retrieved from matching with the exit wave.

In practice it is possible to retrieve the positions of the columns with high accuracy (0.01 nm) and to obtain a rough estimate of the density of the columns. Figure 22 shows a map of the projected potential $Ba_2NaNb_5O_{15}$ retrieved from the exit wave of Fig. 21. Here all atoms are imaged as white dots with an intensity roughly proportional to the weight of the columns.

In principle, the three-dimensional structure can be retrieved by combining the information from different zone orientations. However, the number of 'visible' zone orientations is limited by the resolution of the electron microscope.

Intrinsic Limitations

It should be noticed that HREM, even combined with quantitative reconstruction methods, has its intrinsic limitations.

Although the positions of the projected atom columns can be determined with high accuracy (0.01 nm), the technique is less sensitive for determining the mass density of the columns and to get information about the bonds between atoms. Besides,

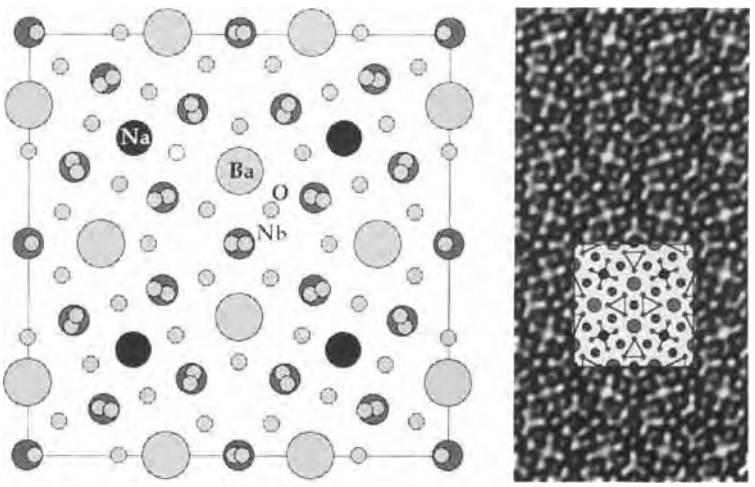


Figure 22. Experimentally retrieved structure for $Ba_2NaNb_5O_{15}$.

due to the high speed of the electrons, they only sense a projected potential so that no information can be obtained about the distribution of this potential along the columns. Three-dimensional information can be obtained though, by investigating the same object along different zone axes. Furthermore, as shown above, for some object thicknesses, atom columns can become extinct so that they cannot be retrieved from the exit wave.

1.1.2.7 Case Studies

Alloys

As shown in Sec. 1.1.2.6, substitutional alloys with a column structure are particularly interesting subjects for HREM studies. If the alloys are oriented with the columns parallel to the electron beam, one can directly visualize the minority columns by imaging without the reflections of the basic f.c.c. structure. The diffraction pattern and the beam selecting aperture are shown in Fig. 23.

Figure 24 shows a historical experimental high-resolution micrograph of Au_4Mn viewed along the (1 0 0) zone using this dark field superlattice imaging mode

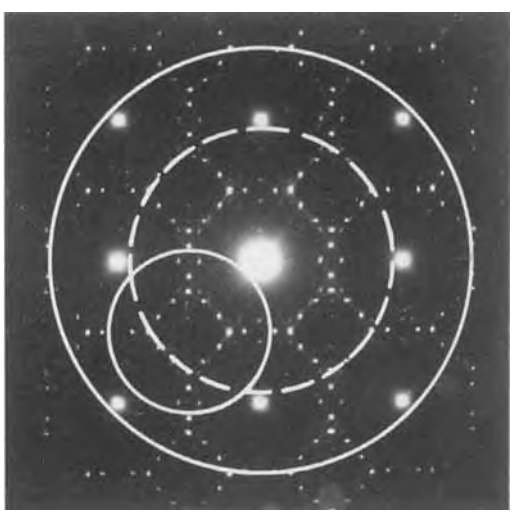


Figure 23. Diffraction pattern and beam selecting apertures for f.c.c.-based substitutional alloys. For high-resolution dark field imaging, the f.c.c. reflections are excluded and only the superstructure reflections of one reciprocal unit cell are included.

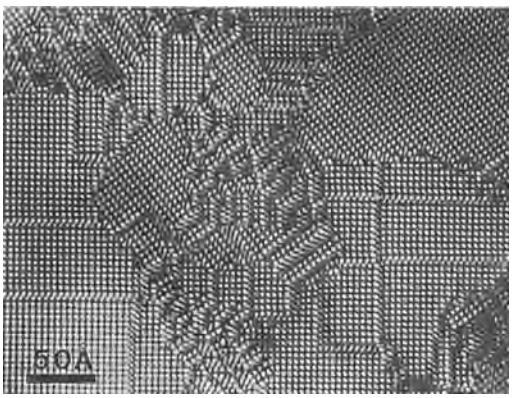


Figure 24. Dark field superlattice image of the alloy Au_4Mn viewed along the column direction [0 0 1]. The manganese columns are selectively imaged as bright dots [9].

using a 200 keV electron microscope with only 0.25 nm point resolution. The bright dots correspond to the configuration of manganese columns. Two orientation variants are present, as well as a number of antiphase boundaries.

The imaging characteristics in which the manganese atoms are revealed as bright dots are preserved even close to the interfaces. This allows deduction of the displacement vectors and the orientations directly from the images.

Figure 25 shows a large period antiphase boundary superstructure 2223 (Fig. 25b). Discommensurations can be seen in a dark field line resolution image (Fig. 25c). Figure 26 shows different models of superstructures in the alloy Nb_5Ga_{13}. Figure 27 shows high-resolution images of Nb_5Ga_{13} viewed along [1 1 0]. From the high-resolution images it can be concluded that the actual structure is that corresponding to the model in Fig. 26a.

Crystal Defects

Grain Boundaries

In non-coherent interfaces the atom columns can still be imaged by white dots but the exact positions of the dots do not necessarily correspond with the exact positions of the columns so that comparison with simulated images is necessary. An example is given in Fig. 28 showing a (3 1 0) $\Sigma = 5$ grain boundary in germanium viewed along [0 0 1] which contains many dislocations. In agreement with the

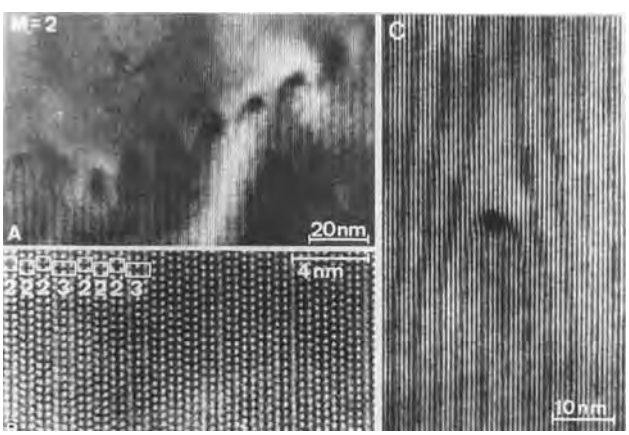


Figure 25. [1 0 0] zone of the f.c.c.-based ordered alloy Au_4Mn [20]. The long period antiphase boundary superstructure 2223 is revealed in the high-resolution image (b). A dark field line resolution image is used to reveal the discommensurations (i.e., the 3-strips).

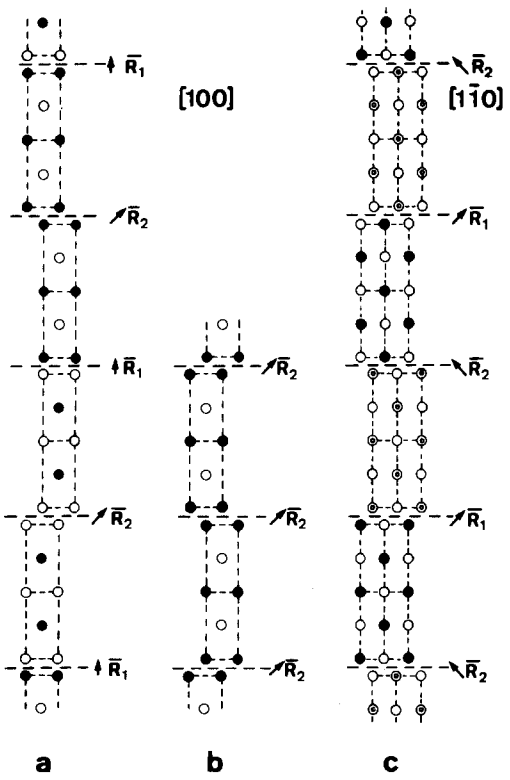


Figure 26. Models of very long period superstructures in the alloy Nb_5Ga_{13}; it contains two types of antiphase boundaries [21].

channeling theory (see Appendix D) the intensity at the column positions varies periodically with depth. Figure 28a shows an image revealing black columns. Figure 28b shows white columns. The point resolution of the microscope in this case is about 0.15 nm (400 keV). If possible it is preferable to search for a thickness for which the columns are imaged as white dots.

Stacking Faults

A stacking fault tetrahedron (SFT) is a defect which is limited by (1 1 1)

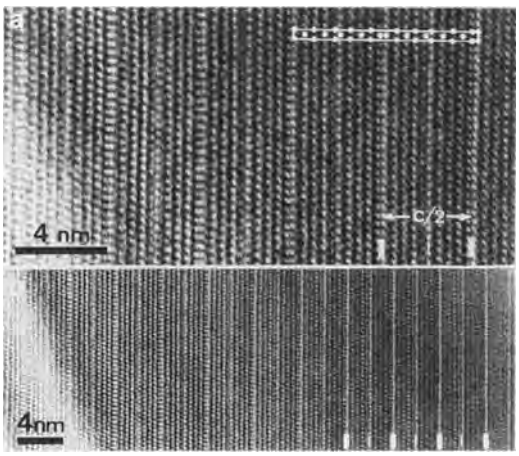


Figure 27. High-resolution images of Nb_5Ga_{13} viewed along the [1 1 0] zone [21]. The types of antiphase boundaries as well as their stacking sequences can be read directly from the image. The actual structure corresponds to the model in Fig. 26a [21].

stacking fault planes intersecting stair red dislocations, and have been observed in low stacking fault energy metals and alloys.

In ion-implanted and annealed silicon, the sizes of the SFTs are suitable to be studied the HREM. When the images are taken with the incident beam along the [1 1 0] zone axis, the images show a 'V'-shaped discontinuity in the rows of bright dots. Within the 'V' zone, the dots are displaced due to the fact that the atom columns are intersected by two stacking faults. The displacement of the dots is related to the length of the displaced column, that is, the displacement is largest near the point of the 'V' and dissipates gradually with increasing distance from this point. Image calculations were performed with the real space method for tetrahedra of different sizes using atom positions derived from models in the literature, for the vacancy type of

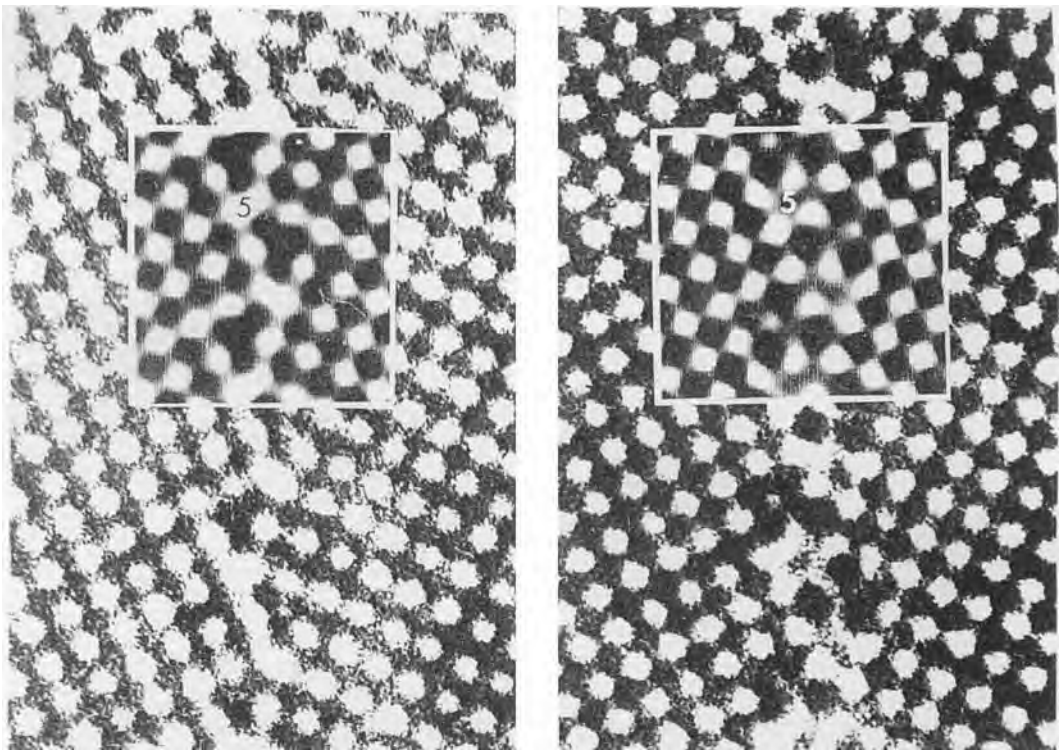


Figure 28. $\Sigma = 5$ grain boundary in germanium. (Courtesy of J. L. Rouvière.)

tetrahedron as well as for the interstitial type. Approximately 10^5 atoms are involved in the calculations. By carefully looking along a glancing incidence (Fig. 29), it is clear that the displacement of the bright dots in the interstitial type is directed towards the point, whereas in the vacancy type the displacement is in the opposite sense, in agreement with experiment. Hence it can be concluded that the stacking fault tetrahedra in silicon are of the vacancy type.

This is a very fortunate case where only two different atomic models are possible which are so different that discrimination can be made on the basis of only a few image simulations.

Dislocations

When edge dislocations are parallel to the zone axis of observation and when the resolution of the microscope is sufficient to discriminate the individual atoms the dislocation structure can be unravelled. Figure 30a represents the high-resolution image of a 60° dissociated dislocation in silicon. The dislocation is dissociated in two Schockley partials, one 90°D and one 30°D enclosing an intrinsic stacking fault. From the high-resolution image one can estimate the dissociation energy. The main feature is that the partials consist of well-defined structural units (Fig. 30b) that also occur in the dislocation cores. Figure

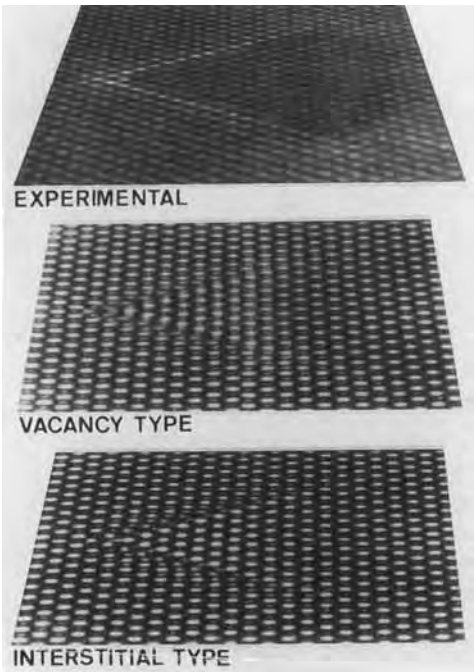


Figure 29. SFT in silicon, viewed under glancing angle. Top: experimental image. Middle: simulated image for a vacancy-type SFT. Bottom: simulated image for an interstitial-type SFT. From this is can be concluded that the SFT is of the vacancy type. (Courtesy of W. Coene and H. Bender.)

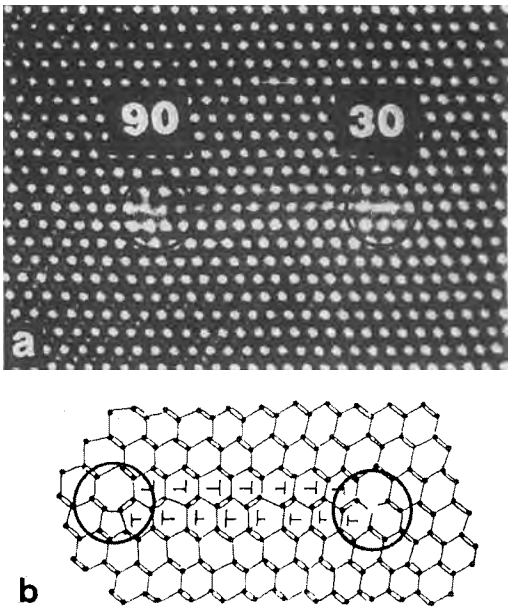


Figure 30. A 60° dissociated dislocation in silicon. (Courtesy of J. Thibault-Desseaux.)

30c shows the dislocation structures occurring at a $\Sigma = 9$ grain boundary in silicon. Here also the same structural units can be observed.

Figure 31 shows an image of a 60° dislocation in silicon dissociated into a 30° and a 90° partial. In Fig. 31 one can observe, in between the partials, different sets of 0.33 nm fringes. What is particularly interesting is that these fringes are caused by $(\bar{4}\,2\,3)/3$ reflections that are forbidden in the bulk crystal. Hence they reveal detailed information about the dislocation cores at the subnanometer level. For details the reader is referred to Alexander et al. [22].

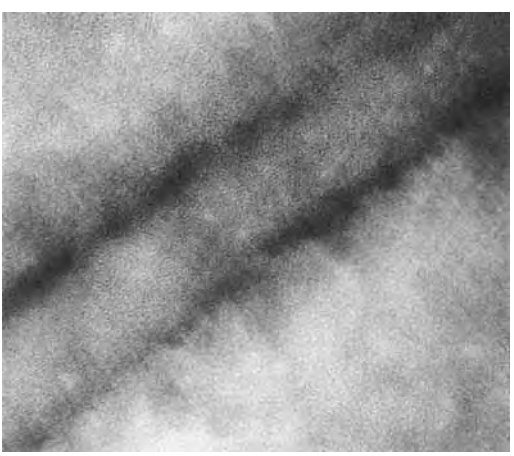


Figure 31. Image of a dissociated 60° dislocation in silicon. In between the 30° and 90° partials, different sets of 0.33 nm fringes are visible. These fringes are caused by forbidden $(\bar{4}\,2\,2)/3$ reflections and contain high-resolution information about the dislocation cores. (Courtesy of J. C. H. Spence.)

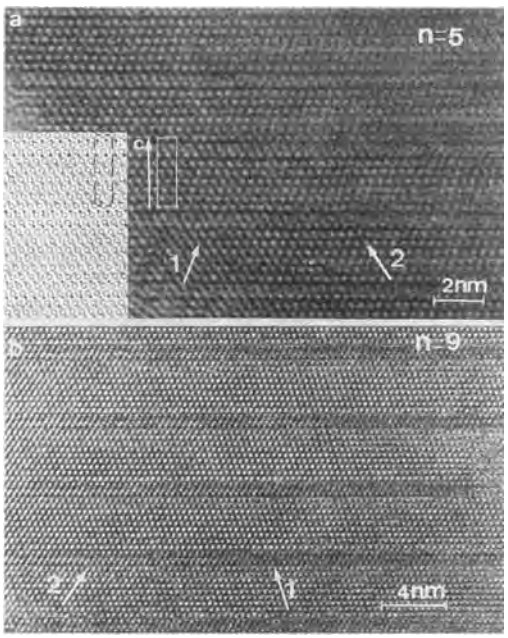


Figure 32. Mixed-layer compounds $As_2Te_3(GeTe)_n$ with (a) $n = 5$ and (b) $n = 9$ [23]. The image for $n = 5$ is compared with a structure model in the inset. The As_2Te_3 layer can clearly be distinguished from the (GeTe) block.

Mixed-Layer Compounds

Mixed-layer compounds can generally be defined as an alternation of stackings of blocks of different types in which the number of blocks and the stacking sequence can be altered by changes in composition and so forth. In a sense, a system with periodical translation interfaces or twin planes can also be considered as a mixed-layer compound. In the literature a more restricted definition can be found in which each type of the basic block has to occur as an existing member of the series.

Figure 32 reveals a mixed layer compound of the series $As_2Te_3(GeTe)_n$ with $n = 5$ (Fig. 32a) and $n = 9$ (Fig. 32b). The systems consist of a regular stacking of As_2Te_3 layers with (GeTe) blocks. A structure model is shown in the inset.

Figure 33 shows an example of an incommensurate stacking in the mixed-layer compound $SnTaS_3$. This is a compound of the type MTS_3, which in this case consists of a stacking of hexagonal TaS_2 and tetragonal SnS layers. The image simulations which confirm the model are shown in Fig. 34. Figure 35 shows the images of a homologous series of polytypoids of the type $La_nTi_{n-\delta}O_{3n}$. The layer sequences can be deduced directly from the images.

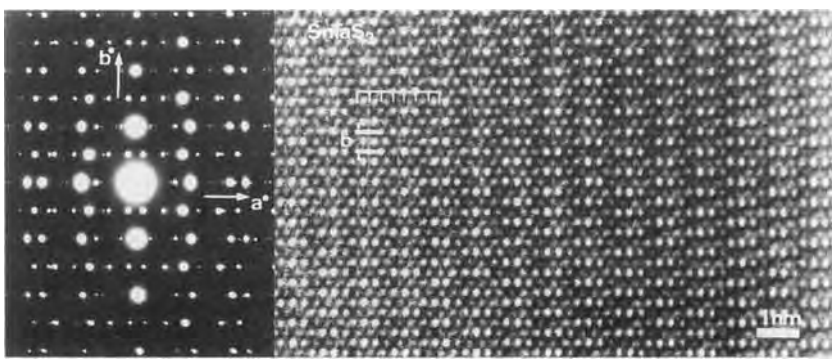


Figure 33. Incommensurate mixed layer compound of the type $MTS_3(SnTaS_3)$ consisting of the juxtaposition of hexagonal TaS_2 and tetragonal SnS layers [24]. (a) The diffraction pattern is the superposition of the diffraction patterns of the separate layers, supplemented with double diffraction spots. (b) High-resolution image along the normal to the layer plane.

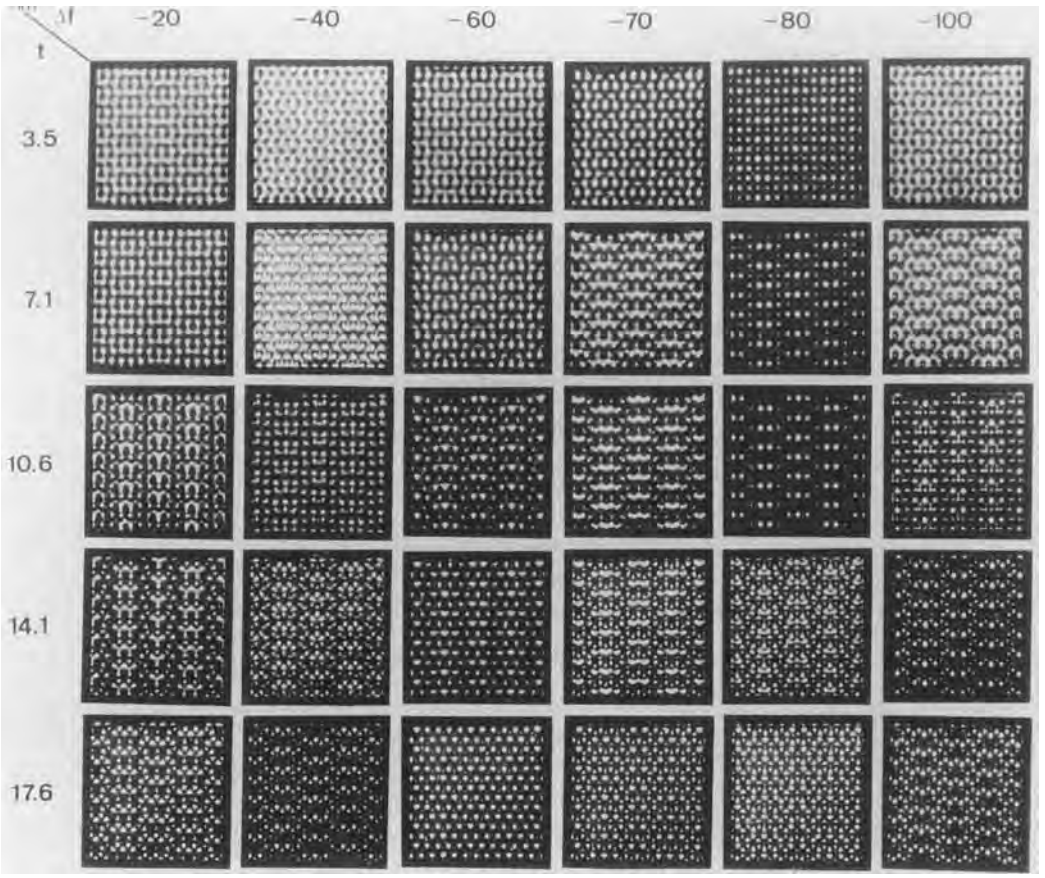


Figure 34. Image simulation corresponding with the experimental situation of Fig. 33 [24]. Images are simulated for various focus values Δf and crystal thickness H (in nm). The best match is obtained for $\Delta f = -60$ nm and $t = 10.6$ nm.

High-T_c Superconductors

Figure 36 shows a high-resolution image of superconducting BiSrCaO viewed along $[1\,0\,0]_p$. From this it can be seen that the cleavage of the crystal has taken place in between two adjacent BiO layers.

Figure 37 shows a high-resolution image of superconducting $YBa_2Cu_3O_{7-\delta}$. $(CuO)_2$ double layers as well as CuO single layers can be identified. Cleavage has taken place at the CuO layer.

From these images it is clear that if the electron beam is nearly parallel to the surfaces of the crystal, the profile of the surface can be visualized directly. This enables one to characterize the surface layers and to detect surface relaxation or modification effects.

Minerals

Figure 38 shows the image of the mineral hollandite ($Ba_xMn_8O_{16}$) viewed along

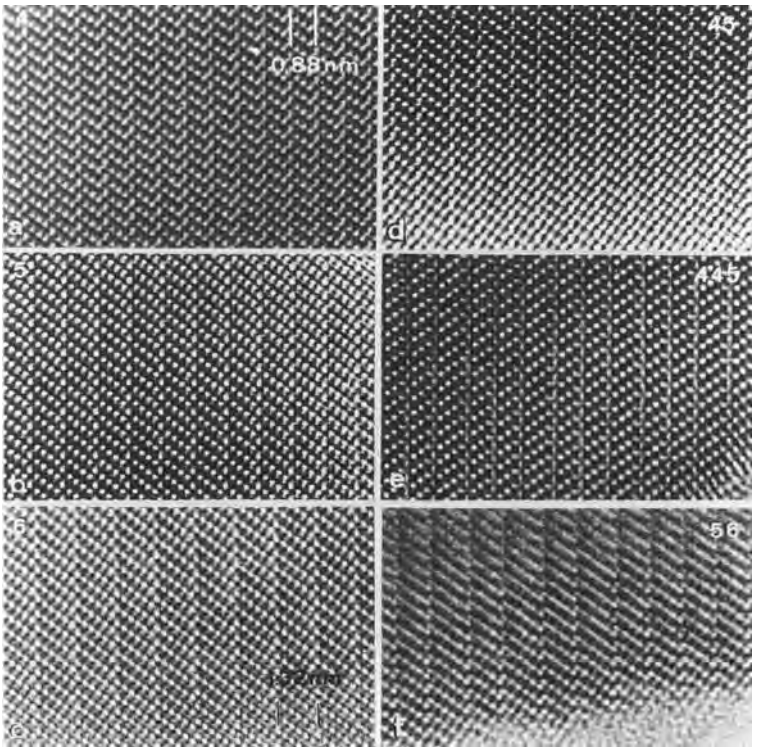


Figure 35. High-resolution image along the [1 1 $\bar{2}$ 0]* zone of a homologous series of polytypoids with composition $La_nTi_{n-\delta}O_{3n}$ [25]. The layer sequences in the different polytypoids, as deduced from the image and from the diffraction pattern, are indicated.

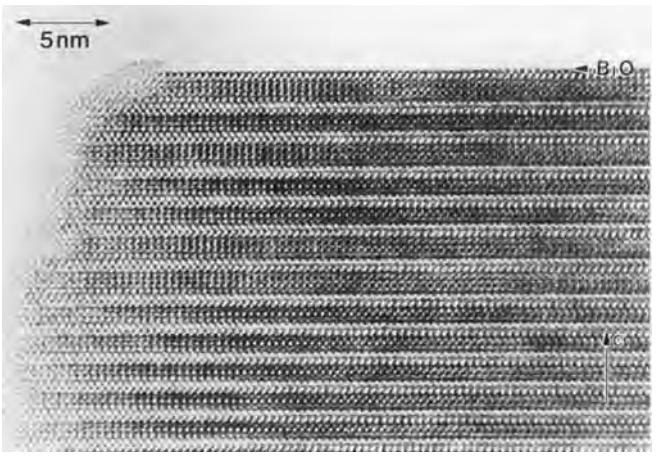


Figure 36. Superconducting BiSrCaO compound exhibiting a cleaved surface [26]. The high-resolution image along $[1\,0\,0]_p$ proves that cleavage took place between the two adjacent BiO layers in the structure.

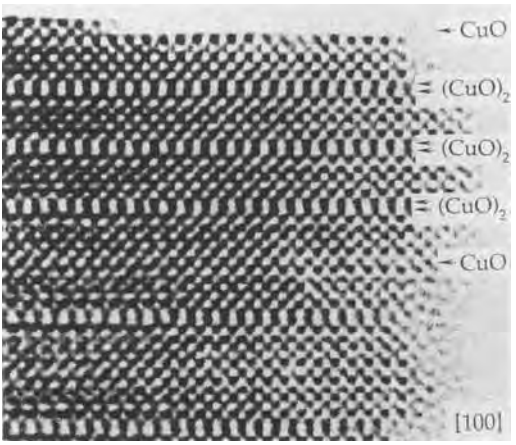


Figure 37. Superconducting $YBa_2Cu_3O_4$. The $(CuO)_2$ double layer as well as (CuO) single layers can be identified. Cleavage has taken place at the CuO layer [27].

[0 1 0]. In this case the images can be interpreted intuitively in terms of the projected potential (see Sec. 1.1.2.6). As can be judged by comparing with the structure model in Fig. 39, a planar defect can be identified. The image matches with the computer simulation (inset).

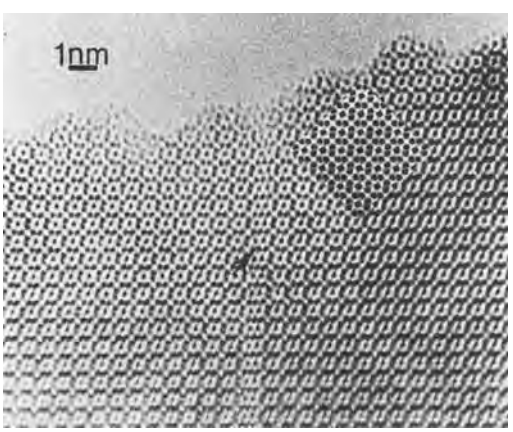


Figure 38. [0 1 0] zone image of the mineral hollandite ($Ba_xMn_8O_{16}$). The barium ions in the channels of the MnO_2 framework are revealed as dark dots. A planar defect consisting of double channels is present. In the top right corner a computer simulated image is superposed on the actual image [28].

Carbon Structures

Figure 40 shows the image of a C_{60} crystal ('bucky balls') containing intersecting intrinsic stacking faults. A model for the intersection is shown in the inset.

Figure 41 shows an image of a rubidium intercalated C_{60} crystal (Rb_6C_{60}). Also in this case the contrast can be explained intuitively in terms of the projected structure (Fig. 42). Figure 43 shows a high-resolution image of straight and helix-shaped nanotubes in which the (0 0 0 1) graphite planes are resolved.

Cage Structures

Figures 44 and 45 show images of the cage structure of Na_8Si_{46} and Na_xSi_{136} viewed along (0 0 1), respectively (0 1 1). Since these structures are very open, the images can be directly interpreted in terms of the projected structure viewed along these directions (Fig. 46, respectively Fig. 47).

1.1.2.8 Appendices

Appendix A: Image Formation [1]

The wavefunction in the image plane is given by

$$\phi(\boldsymbol{R}) = F_{\boldsymbol{R}}^{-1} A(\boldsymbol{g}) \exp[-i\chi(\boldsymbol{g})] \times D(\alpha, \Delta, \boldsymbol{g}) F_{\boldsymbol{g}} \psi(\boldsymbol{R}) \qquad (36)$$

$A(\boldsymbol{g})$ represents the physical aperture with radius g_A selecting the imaging beams, thus

$$A(\boldsymbol{g}) = \begin{cases} 1 & \text{for } |g| \leqslant g_A \\ 0 & \text{for } |g| > g_A \end{cases}$$

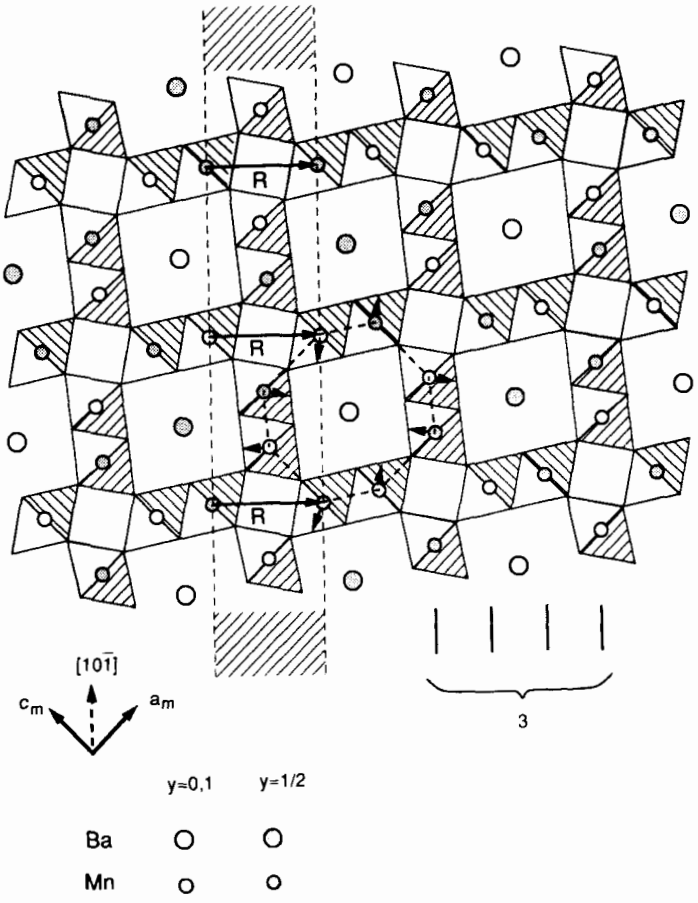


Figure 39. Schematic representation of the hollandite structure along the [0 1 0] zone. A double row planar defect as seen in Fig. 38 is generated by removing the slice indicated by dotted lines and closing the gap [28].

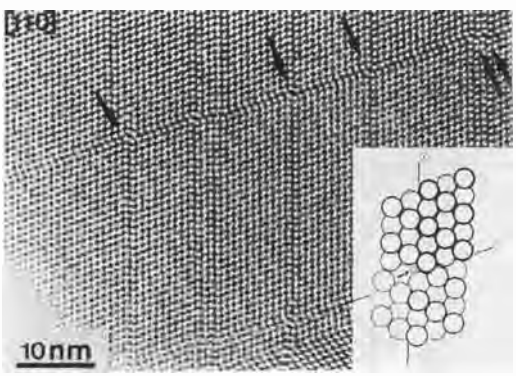


Figure 40. [1 1 0] zone image of a C_{60} crystal containing intersecting intrinsic stacking faults. Along the intersection line a row of partial vacancies is formed as represented schematically in the inset [29].

The total phase shift due to spherical aberration and defocus is

$$\chi(\boldsymbol{g}) = \tfrac{1}{2}\pi C_s \lambda^3 g^4 + \pi\epsilon\lambda g^2 \qquad (37)$$

with C_s the spherical aberration coefficient, ϵ the defocus, and λ the wavelength. The imaging process is also influenced by spatial and temporal incoherence effects. Spatial incoherence is caused by the fact that the illuminating beam is not parallel but can be considered as a cone of incoherent plane waves (beam convergence). The image then results from a superposition of the respective image intensities. Temporal incoherence results

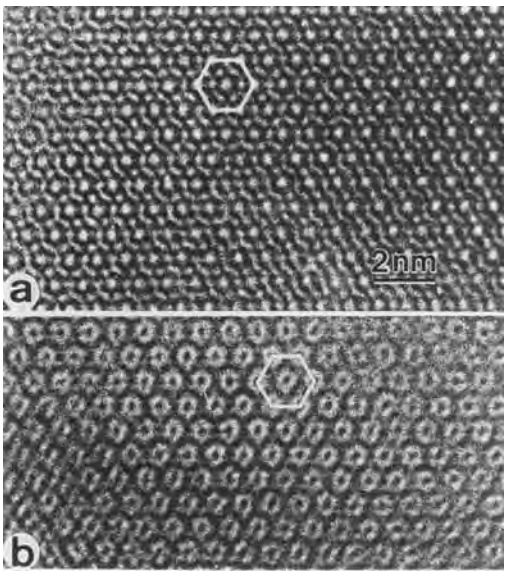


Figure 41. (a) Rubidium intercalated C_{60} crystal (Rb_6C_{60}). The circular bright areas image the C_{60} molecules: the dark areas represent the Rb ions. The structure is imaged along the [1 1 1] zone; (b) different defocus [30].

from fluctuations in the energy of the thermally emitted electrons, in the fluctuation of the lens currents, and of the accelerating voltage. All these effects cause the focus ϵ to fluctuate. The final image is then

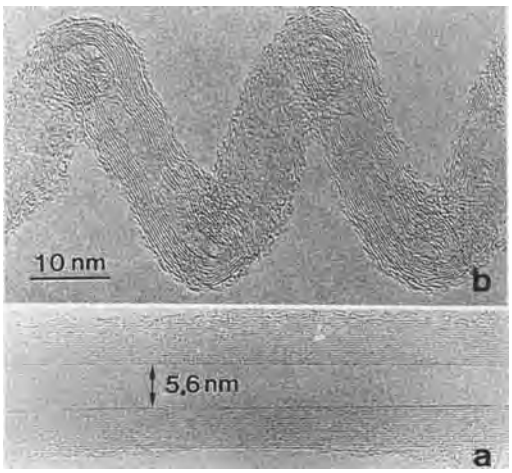


Figure 43. High-resolution images of carbon nanotubes [31]. (a) Straight tube; only the (0 0 0 1) planes are resolved. (b) Helix-shaped nanotubes; (0 0 0 1) planes are resolved.

the superposition (integration) of the images corresponding with the different incident beam directions $\boldsymbol{K}$ and focus values ϵ, that is,

$$I(\boldsymbol{R}) = \int_{\epsilon}\int_{\boldsymbol{K}} |\phi(\boldsymbol{R}, \boldsymbol{K}, \epsilon)|^2 f_s(\boldsymbol{K}) f_T(\epsilon)\, d\boldsymbol{K}\, d\epsilon \tag{38}$$

where $\phi(\boldsymbol{R}, \boldsymbol{K}, \epsilon)$ denotes that the

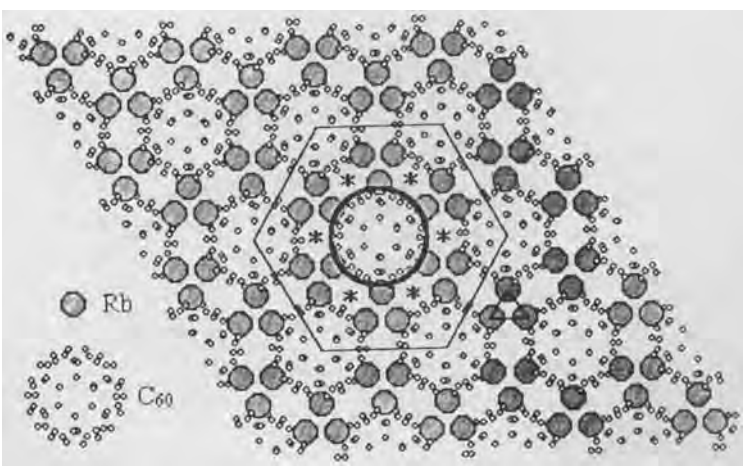


Figure 42. Structure model, projected along the [1 1 1] of the intercalate Rb_6C_{60}. This model can be compared with the image of Fig. 41b [30].

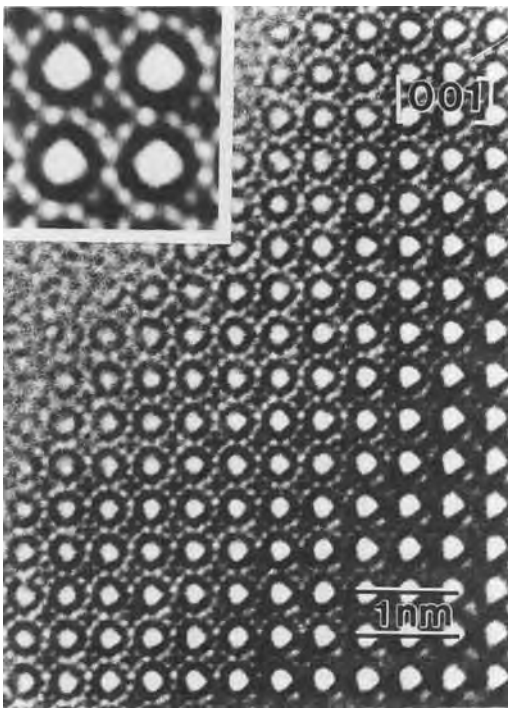


Figure 44. High-resolution image of the cage structure Na_8Si_{46} [32]. The bright dots correspond to open channels in the structure. The inset is an averaged image.

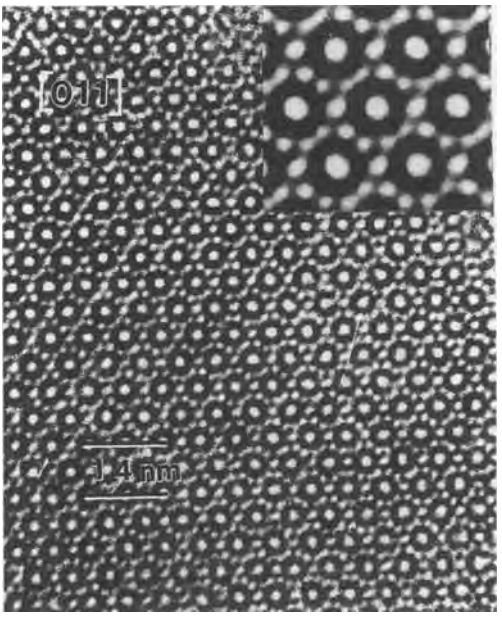


Figure 45. High-resolution image of the cage structure Na_xSi_{136} [32]. The bright dots correspond to open channels in the structure. The inset is an averaged image.

wavefunction in the image plane also depends on the incident wavevector $\boldsymbol{K}$ and on the defocus ϵ. $f_s(\boldsymbol{K})$ and $f_c(\epsilon)$ are the probability distribution functions of $\boldsymbol{K}$, respectively ϵ. Expressions (36), (37), and (38) are the basic expressions describing the whole real imaging process. They are also used for the computer simulation of high-resolution images. However, the computation of Eq. (38) requires the computation of $\psi(\boldsymbol{R})$ for a large number of defocus values and beam directions, which in practice is a horrible task. For this reason, Eq. (38) has often been approximated. In order to study the effect of chromatic aberration and beam convergence (on a more intuitive basis) we assume a disk-like effective source function

$$f_s(\boldsymbol{K}) = \begin{cases} 1 & \text{for } |\boldsymbol{K}| < \alpha/\lambda \\ 0 & \text{for } |\boldsymbol{K}| > \alpha/\lambda \end{cases}$$

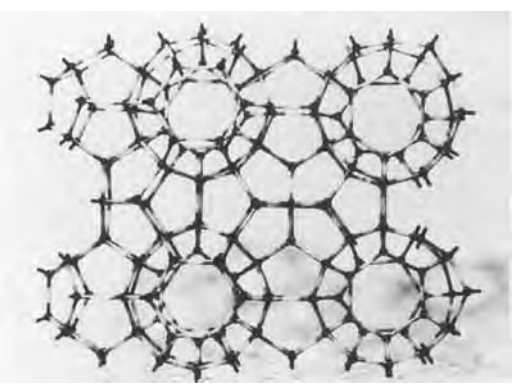


Figure 46. Model for the structure of Na_8Si_{46} as viewed along the same zone as the image of Fig. 44 [32].

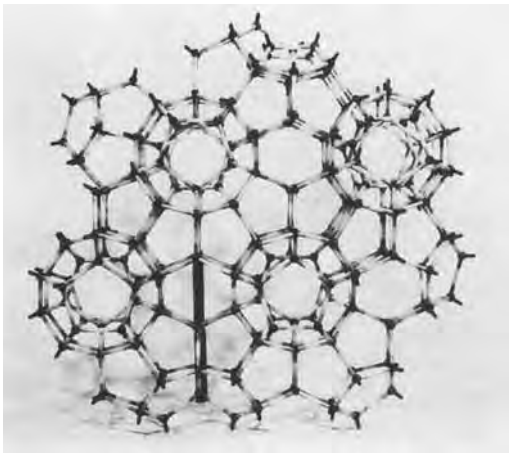


Figure 47. Model for the cage structure Na_xSi_{136} which can be compared with the image of Fig. 45 [32].

with α the apex angle of the illumination cone. We assume further that the integrations over defocus and beam convergence can be performed coherently, that is, over the amplitudes rather than the intensities. This latter assumption is justified when the intensity of the central beam is much larger than the intensities of the diffracted beams so that cross-products between diffracted beam amplitudes can be neglected. We assume that the defocus spread $f_T(\epsilon)$ is a Gaussian centered on ϵ with a half-width Δ. Assuming the object function $\psi(\boldsymbol{R})$ to be independent of the inclination $\boldsymbol{K}$, which is only valid for thin objects, one then finally finds that the effect of the chromatic aberration, combined with beam convergence, can be incorporated by multiplying the transfer function with an effective aperture function

$$D(\alpha, \Delta, \boldsymbol{g}) = B(\Delta, \boldsymbol{g})C(\alpha, \Delta, \boldsymbol{g}) \tag{39}$$

where

$$B(\Delta, \boldsymbol{g}) = \exp(-\tfrac{1}{2}\pi^2\lambda^2\Delta^2 g^4) \tag{40}$$

representing the effect of the defocus spread, and

$$C(\alpha, \Delta, \boldsymbol{g}) = 2J_1(|\boldsymbol{q}|)/|\boldsymbol{q}| \tag{41}$$

with J_1 the Bessel function and $|\boldsymbol{q}| = (\boldsymbol{q}\cdot\boldsymbol{q})^{1/2}$, which may be a complex function for a complex $\boldsymbol{q}$

$$\boldsymbol{q} = 2\pi\alpha\boldsymbol{g}[\epsilon + \lambda g^2(\lambda C_s - i\pi\Delta^2)] \tag{42}$$

$C(\alpha, \Delta, \boldsymbol{g})$ represents the combined effect of beam convergence and defocus spread.

The total image transfer can now be described from Eqs. (5) and (6) as

$$\phi(\boldsymbol{R}) = F_{\boldsymbol{R}}^{-1}A(\boldsymbol{g})\exp[-i\chi(\boldsymbol{g})] \times D(\alpha, \Delta, \boldsymbol{g})F_{\boldsymbol{g}}\psi(\boldsymbol{R}) \tag{43}$$

that is, the effective aperture yields a damping envelope function for the phase transfer function. Other approximations for including the effects of beam convergence and chromatic aberrations [33] using a Gaussian effective source lead to a similar damping envelope function. Experimentally obtained transfer functions confirm this behavior.

In Eq. (43) the incoherent effects are approximated by a coherent envelope function. Hence it is called the coherent approximation. It is usually valid for thin objects. A full treatment of incoherent effects requires the calculation of the double integral in Eq. (38). Another approximation which is valid for thicker objects is based on the concept of the TCC [34]. Here it is assumed that beam convergence and defocus spread do not influence the diffraction in the object. Hence in Eq. (5) they do not appear in the object wavefunction but only in the phase transfer function. Now the wavefunction in the image plane (Eq. (5)) can be written as

$$\Phi(\boldsymbol{R}, \boldsymbol{K}, \epsilon) = F_{\boldsymbol{R}}^{-1}T(\boldsymbol{g}, \boldsymbol{K}, \epsilon)\psi(\boldsymbol{g}) \tag{45}$$

with

$$T(\boldsymbol{g}, \boldsymbol{K}, \epsilon) = A(\boldsymbol{g}) \exp[-\mathrm{i}\chi(\boldsymbol{g}, \boldsymbol{K}, \epsilon)] \quad (46)$$

Substituting Eq. (8) into Eq. (38) then yields, after Fourier transforming,

$$I(\boldsymbol{g}) = F_{\boldsymbol{g}}[I(\boldsymbol{R})] = \int \psi(\boldsymbol{g}+\boldsymbol{g}')\tau(\boldsymbol{g}+\boldsymbol{g}', \boldsymbol{g}')\psi^*(\boldsymbol{g}')\,\mathrm{d}\boldsymbol{g}' \quad (47)$$

with

$$\tau(\boldsymbol{g}+\boldsymbol{g}', \boldsymbol{g}') = \int_\epsilon \int T^*(\boldsymbol{g}+\boldsymbol{g}', \boldsymbol{K}, \epsilon) \times T(\boldsymbol{g}', \boldsymbol{K}, \epsilon)\,\mathrm{d}\boldsymbol{K}\,\mathrm{d}\epsilon \quad (48)$$

τ is the transmission cross-coefficient. It describes how the beams $\boldsymbol{g}'$ and $\boldsymbol{g}+\boldsymbol{g}'$ are coupled to yield the Fourier component $\boldsymbol{g}$ of the image intensity.

Appendix B: Scattering of an Electron in a Thin Object

We will now follow a classical approach.

The non-relativistic expression for the wavelength of an electron accelerated by an electrostatic potential E is given by

$$\lambda = \frac{h}{\sqrt{2meE}} \quad (49)$$

with h the Planck constant, m the electron mass, and e the electron charge.

During the motion through an object with local potential $V(x,y,z)$ the wavelength will vary with the position of the electron as

$$\lambda'(x,y,z) = \frac{h}{\sqrt{2me[E+V(x,y,z)]}} \quad (50)$$

For thin phase objects and large accelerating potentials the assumption can be made that the electron keeps traveling along the z direction so that by propagation through a slice $\mathrm{d}z$ the electron suffers a phase shift:

$$\begin{aligned}\mathrm{d}\chi(x,y,z) &= 2\pi\frac{\mathrm{d}z}{\lambda'} - 2\pi\frac{\mathrm{d}z}{\lambda} \\ &= 2\pi\frac{\mathrm{d}z}{\lambda}\left(\frac{\sqrt{E+V(x,y,z)}}{\sqrt{E}} - 1\right) \\ &\simeq \sigma V(x,y,z)\,\mathrm{d}z \end{aligned} \quad (51)$$

with

$$\sigma = \pi/\lambda E$$

so that the total phase shift is given by

$$\chi(x,y) = \sigma \int V(x,y,z)\,\mathrm{d}z = \sigma V_{\mathrm{p}}(x,y) \quad (52)$$

where $V_{\mathrm{p}}(x,y)$ represents the potential of the specimen projected along the z direction.

Under this assumption the specimen acts as a pure phase object with transmission function

$$\psi(x,y) = \exp[\mathrm{i}\sigma V_{\mathrm{p}}(x,y)] \quad (53)$$

When the object is very thin, one has

$$\psi(x,y) \approx 1 + \mathrm{i}\sigma V_{\mathrm{p}}(x,y) \quad (54)$$

This is the weak phase approximation.

The effect of all processes, prohibiting the electrons from contributing to the image contrast, including the use of a finite aperture can in a first approximation be represented by a projected absorption function in the exponent of Eq. (53), so that

$$\psi(x,y) = \exp[\mathrm{i}\sigma V_{\mathrm{p}}(x,y) - \mu(x,y)] \quad (55)$$

or

$$\psi(\boldsymbol{R}) = \exp[\mathrm{i}\sigma V_{\mathrm{p}}(\boldsymbol{R}) - \mu(\boldsymbol{R})] \quad (56)$$

with $\boldsymbol{R} = (x,y)$ the vector in the plane perpendicular to z.

Appendix C: Scattering of an Electron in a Thick Object [35]

Optical Approach: The Multislice Method [36]

Although the multislice formula can be derived from quantum mechanical principles, we follow a simplified version of the more intuitive original optical approach. A more rigorous treatment is given in the next section.

Consider a plane wave, incident on a thin specimen foil and nearly perpendicular to the incident beam direction z. If the specimen is sufficiently thin, we can assume the electron to move approximately parallel to z so that the specimen acts as a pure phase object with transmission function Eq. (53):

$$\psi(x,y) = \exp[i\sigma V_p(x,y)]$$

A thick specimen can now be subdivided into thin slices, perpendicular to the incident beam direction. The potential of each slice is projected into a plane which acts as a two-dimensional phase object. Each point (x,y) of the exit plane of the first slice can be considered as a Huyghens source for a secondary spherical wave with amplitude $\psi(x,y)$ (Fig. 48).

Now the amplitude $\psi(x',y')$ at the point (x',y') of the next slice can be found by the superposition of all spherical waves of the first slice, that is, by integration over x and y, yielding

$$\psi(x',y') = \int \exp[i\sigma V_p(x,y)] \times \frac{\exp(2\pi i k r)}{r}\,dx\,dy$$

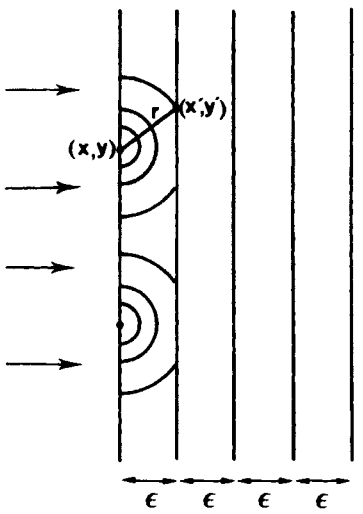


Figure 48. Schematic representation of the propagation effect of electrons between successive slices of thickness ϵ.

When $|x-x'| \ll \epsilon$ $|y-y'| \ll \epsilon$, with ϵ the slice thickness, the Fresnel approximation can be used, that is,

$$r = \sqrt{(x-x')^2 + (y-y')^2 + \epsilon^2} \approx \epsilon\left(1 + \frac{(x-x')^2}{2\epsilon^2} + \frac{(y-y')^2}{2\epsilon^2}\right)$$

so that

$$\psi(x',y') \approx \frac{\exp(2\pi i k\epsilon)}{\epsilon} \times \int \exp[i\sigma V_p(x,y)] \times \exp\left[i\frac{\pi k}{\epsilon}[(x-x')^2 + (y-y')^2]\right] dx\,dy$$

which, apart from constant factors, can be written as a convolution product:

$$\psi(x,y) = \epsilon[i\sigma V_p(x,y)] * \exp[i\pi k(x^2+y^2)/\epsilon] \qquad (57)$$

where the convolution product of two functions is defined as (in one dimension)

$$f(x) * g(x) = \int f(x')g(x - x')\,\mathrm{d}x'$$

If the wavefunction at the entrance face is $\psi(x, y, 0)$, instead of a plane wave one has for the wavefunction at the exit face

$$\psi(x, y, \epsilon) = \{\psi(x, y, 0) \exp[\mathrm{i}\sigma V_\mathrm{p}(x, y)]\} * \exp[\mathrm{i}\pi k(x^2 + y^2)/\epsilon] \qquad (58)$$

This is the Fresnel approximation in which the emerging spherical wavefront is approximated by a paraboloidal wavefront.

The propagation through the vacuum gap from one slice to the next is thus described by a convolution product in which each point source of the previous slice contributes to the wavefunction in each point of the next slice. The motion of an electron through the whole specimen can now be described by an alternating of phase object transmissions (multiplications) and vacuum propagations (convolutions). In the limit of the slice thickness ϵ tending to zero, this multislice expression converges to the exact solution of the non-relativistic Schrödinger equation in the forward-scattering approximation.

In the original multislice method one used the Fourier transform of Eq. (58) where the real space points (x, y) are transformed into diffracted beams $\boldsymbol{g}$ and where convolution and normal products are interchanged, that is,

$$\psi(\boldsymbol{g}, \epsilon) = \{\psi(\boldsymbol{g}, 0) * \exp[\mathrm{i}\sigma V_{\boldsymbol{g}}]\} \times \exp[\mathrm{i}\pi g^2 \epsilon/k] \qquad (59)$$

where $V_{\boldsymbol{g}}$ are the structure factors (Fourier transforms of the unit cell potential).

The wavefunction at the exit face of the crystal can now be obtained by successive application of Eq. (58) or (59). This can either be done in real space (Eq. (58)) or in reciprocal space (Eq. (59)). The major part of the computing time is required for the calculation of the convolution product, which is proportional to N^2 (N is the number of sampling points (real space) or beams (reciprocal space)).

Since the Fourier transform of a convolution product yields a normal product (with calculation time proportional to N), a large gain in speed can be obtained by alternatively performing the propagation in reciprocal space and the phase object transmission in real space [37]. In this way the computing time is devoted to the Fourier transforms and is proportional to $N \log_2 N$.

Another way of increasing the speed is in the so-called real space method [38]. Here the whole calculation is done in real space using Eq. (58) but the forward scattering of the electrons is exploited so as to calculate the convolution effect of the propagation only in a limited number of adjacent sampling points. In this way, the calculation time is proportional to N. This method does not require a periodic crystal and is thus suitable for calculation of crystal defects.

Quantum Mechanical Approach [35]

As is clear from Eq. (5) the calculation of the image wavefunction $\Phi(\boldsymbol{R})$ requires the knowledge of $\psi(\boldsymbol{R})$, that is, the wavefunction at the exit face of the object. This can be obtained by numerically solving the Schrödinger equation in the object. For convenience we will now follow a simplified more intuitive approach, which leads to the correct results.

If we assume that the fast electron, in the direction of propagation (z axis) behaves as a classical particle with velocity $v = hk/m$ we can consider the z axis as a time axis with

$$t = \frac{mz}{hk} \tag{60}$$

Hence we can start from the time-dependent Schrödinger equation

$$-\frac{\hbar}{\mathrm{i}}\frac{\partial\psi}{\partial t}(\bar{R}, t) = H\psi(\bar{R}, t) \tag{61}$$

with

$$H = -\frac{\hbar^2}{2m}\Delta_{\boldsymbol{R}} - eU(\boldsymbol{R}, t) \tag{62}$$

with $U(\boldsymbol{R}, t)$ the electrostatic crystal potential, m and k the relativistic electron mass and wavelength, and $\Delta_{\boldsymbol{R}}$ the Laplacian operator acting in the plane ($\boldsymbol{R}$) perpendicular to z.

Using Eq. (60) we then have

$$\frac{\partial\psi(\boldsymbol{R}, z)}{\partial z} = \frac{\mathrm{i}}{4\pi k}(\Delta_{\boldsymbol{R}} + V(\boldsymbol{R}, z))\psi(\boldsymbol{R}, z) \tag{63}$$

with

$$V(\boldsymbol{R}, z) = \frac{2me}{\hbar^2}U(\boldsymbol{R}, z) \tag{64}$$

This is the well-known high-energy equation in real space which can also be derived from the stationary Schrödinger equation in the forward-scattering approximation [35].

In HREM of crystalline objects, the object is usually oriented along a zone axis, so that the electrons are traveling parallel to the atom columns. If the periodicity along the column direction is not too large (less than 1–2 nm), the fast electron does not feel this variation. In fact it sees the potential as constant along z. In other words, the effect of higher-order Laue zones or upper layer lines is negligible. This is the projection approximation, which is usually valid for most high-resolution conditions, Now Eq. (63) becomes

$$\frac{\partial\psi(\boldsymbol{R}, z)}{\partial z} = \frac{\mathrm{i}}{4\pi k}(\Delta_{\boldsymbol{R}} + V(\boldsymbol{R}))\psi(\boldsymbol{R}, z) \tag{65}$$

with

$$V(\boldsymbol{R}) = \frac{2me}{\hbar}\frac{1}{z}\int_0^z U(\boldsymbol{R}, z)\,\mathrm{d}z \tag{66}$$

the potential, averaged (projected) along z. In the time-dependent Schrödinger picture (Eq. (61)) the electron walks as a function of time in a two-dimensional potential of projected atom columns.

Equation (65) can also be transformed to reciprocal space.

Assuming $V(\boldsymbol{R})$ to be periodic in two dimensions, we can expand it in Fourier series,

$$V(\boldsymbol{R}) = \sum_{\boldsymbol{g}} V_{\boldsymbol{g}} \exp 2\pi\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{R} \tag{67}$$

with $\boldsymbol{g}$ in the zone plane. $V_{\boldsymbol{g}}$ are structure factors. Similarly we have

$$\psi(\boldsymbol{R}) = \sum_{\boldsymbol{g}} \psi_{\boldsymbol{g}}(z) \exp 2\pi\mathrm{i}\boldsymbol{g}\cdot\boldsymbol{R} \tag{68}$$

$\psi_{\boldsymbol{g}}(z)$ represents the amplitude of the beam $\boldsymbol{g}$ at a depth z. Substitution in Eq. (65) then yields

$$\frac{\mathrm{d}\psi_{\boldsymbol{g}}}{\mathrm{d}z} = \mathrm{i}\pi\left[2s_{\boldsymbol{g}}\psi_{\boldsymbol{g}}(z) + \sum_{\boldsymbol{g}'} V_{\boldsymbol{g}-\boldsymbol{g}'}\psi_{\boldsymbol{g}'}(z)\right] \tag{69}$$

with

$$s_{\boldsymbol{g}} = g^2/2k \tag{70}$$

the excitation error, which is approximately equal to the distance between the

reciprocal node $\boldsymbol{g}$ and the Ewald sphere, measured along z. This system of coupled first-order differential equations was derived in the early 1960s (for references, see Van Dyck [35]). Most of the image simulation programs are based on a numerical solution of the dynamical equation in real space (Eq. (65)) or reciprocal space (Eq. (69)), or a combination of both.

The dynamic relation (63) or (65) is a mixture of two equations, each representing a different physical process.

$$\frac{\partial \psi(\boldsymbol{R}, z)}{\partial z} = \frac{\mathrm{i}}{4\pi k} \Delta_{\boldsymbol{R}} \psi(\boldsymbol{R}, z) \tag{71}$$

is a complex diffusion-type of equation, which represent the free electron propagation and whose solution can be represented formally as

$$\psi(\boldsymbol{R}, z) = \exp\left(\frac{\mathrm{i}\Delta_{\boldsymbol{R}} z}{4\pi k}\right) \psi(\boldsymbol{R}, 0) \tag{72}$$

It can also be written as a convolution product:

$$\psi(\boldsymbol{R}, z) = \exp\frac{\mathrm{i}\pi k R^2}{z} * \psi(\boldsymbol{R}, 0) \tag{73}$$

The other part of Eq. (65) is a differential equation,

$$\frac{\partial \psi(\boldsymbol{R}, z)}{\partial z} = \frac{\mathrm{i}}{4\pi k} V(\boldsymbol{R}, z) \psi(\boldsymbol{R}, z) \tag{74}$$

which represents the scattering of the electron by the crystal potential. It can be readily integrated in real space, yielding

$$\psi(\boldsymbol{R}, z) = \exp\left(\frac{\mathrm{i}}{4\pi k} V(\boldsymbol{R}) z\right) \psi(\boldsymbol{R}, 0) \tag{75}$$

with $V(\boldsymbol{R})$ the projected potential as defined in Eq. (66).

The wavefunction in real space is multiplied with a phase factor which is proportional to the electrostatic potential of the object projected along z and which is called the phase object function. The solution of the complete dynamic equation (65) can be written formally as

$$\psi(\boldsymbol{R}, z) \exp\left(\frac{\mathrm{i}}{2\pi k} [\Delta_{\boldsymbol{R}} + V(\boldsymbol{R})] z\right) \psi(\boldsymbol{R}, 0) \tag{76}$$

For the explicit calculation, slice methods are the most appropriate. Here the crystal is cut into thin slices with thickness ϵ perpendicular to the incident beam.

If the slice thickness is sufficiently small, the solution within one slice is approximated by

$$\psi(\boldsymbol{R}, z + \epsilon) = \exp\left(\frac{\mathrm{i}}{2\pi k} \Delta_{\boldsymbol{R}} \epsilon\right) \times \exp\left(\frac{\mathrm{i}}{4\pi k} V(\boldsymbol{R}) \epsilon\right) \psi(\boldsymbol{R}, z)$$

or explicitly

$$\psi(\boldsymbol{R}, z + \epsilon) = \exp\frac{\mathrm{i}\pi k R^2}{\epsilon} * \left[\exp\left(\frac{\mathrm{i}}{4\pi k} V(\boldsymbol{R}) \epsilon\right) \psi(\boldsymbol{R}, 0)\right] \tag{77}$$

This expression is essentially the same as Eq. (58), which was derived from an optical approach. In practice the wavefunction is sampled in a network of closely spaced points. At each point the wavefunction is multiplied with the phase object function. Then the wavefunction is propagated to the next slice, and so on. Calling N the number of sampling points, the phase object requires a calculation time proportional to N^2. In reciprocal space, direct and convolution products are

interchanged, yielding

$$\psi_g(z+\epsilon) = \exp\frac{i\pi g^2\epsilon}{k} \times \left[\mathcal{F}_g\left(\exp\frac{i}{4\pi k}V(\boldsymbol{R})\epsilon\right) * \psi_g(z)\right] \quad (78)$$

This expression is the same as Eq. (59). Now the calculation time of the propagation is proportional to N, the number of beams, whereas the scattering in the phase object gives a calculation time proportional to N^2.

In order to speed up the calculation, the phase object is calculated in real space, and the propagator in reciprocal space [37]. Between each a fast Fourier transform is performed, the calculation time of which is only proportional to $N \log_2 N$. In the standard slice programs the object is assumed to be a perfect crystal. Defects are treated by the periodic continuation method in which the defect is artificially repeated so as to create an artificial supercrystal.

In the real space method, proposed in [38], the whole calculation is performed in real space, but due to the forward scattering of the electrons the propagation effect is limited to a local area so that the calculation time remains proportional to N. This is particularly interesting for treating extended or aperiodic structures.

Appendix D: Electron Channeling

Principle

Although the slice methods are valuable for numerical purposes, they do not provide much physical insight into the diffraction process. There is need for a simple intuitive theory that is valid for larger crystal thicknesses. In our view, a channeling theory fulfills this need. Indeed, it is well known that, when a crystal is viewed along a zone axis, that is, parallel to the atom columns, the high-resolution images often show a one-to-one correspondence with the configuration of columns provided the distance between the columns is large enough and the resolution of the instrument is sufficient. This is for instance the case in ordered alloys with a column structure (see Secs. 1.1.2.6 and 1.1.2.7). From this, it can be suggested that, for a crystal viewed along a zone axis with sufficient separation between the columns, the wavefunction at the exit face mainly depends on the projected structure, that is, on the type of atom columns. Hence, the classical picture of electrons traversing the crystal as plane-like waves in the directions of the Bragg beams which stems from the X-ray diffraction picture and upon which most of the simulation programs are based is, in fact, misleading. The physical reason for this 'local' dynamic diffraction is the channeling of the electrons along the atom columns parallel to the beam direction. Due to the positive electrostatic potential of the atoms, a column acts as a guide or channel for the electron [39, 40] within which the electron can scatter dynamically without leaving the column (Fig. 49).

The importance of channeling for interpreting high-resolution images has often been ignored or underestimated, probably because of the fact that, for historical reasons, dynamic electron diffraction is often described in reciprocal

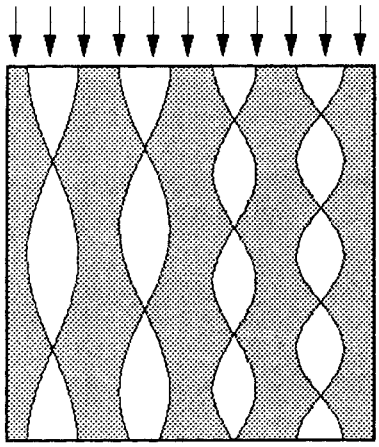


Figure 49. Schematic representation of electron channeling.

space. However, since most of the high-resolution images of crystals are taken in a zone axis orientation, in which the projected structure is the simplest, but in which the number of diffracted beams are the largest, a simple real-space channeling theory yields a much more useful and intuitive, albeit approximate, description of the dynamic diffraction, which allows an intuitive interpretation of high-resolution images, even for thicker objects.

We will now give a simplified version of the basic principles of the theory. For more details the reader is referred to the literature [40].

Isolated Columns

Consider an isolated column of atoms, parallel to the electron beam. If we now consider the depth proportional to the time, the dynamic equation (65) represents the walk of an electron in the two-dimensional projected potential of the columns.

The solution can be expanded in eigenfunctions (eigenstates) of the Hamiltonian

$$\psi(\boldsymbol{R}, z) = \sum_n C_n \phi_n(\boldsymbol{R}) \exp\left(-i\pi \frac{E_n}{E} \frac{z}{\lambda}\right) \tag{79}$$

where

$$H\phi_n(\boldsymbol{R}) = E_n \phi_n(\boldsymbol{R}) \tag{80}$$

with the Hamiltonian

$$H = -\frac{\hbar^2}{2m} \Delta_{\boldsymbol{R}} - eU(\boldsymbol{R}) \tag{81}$$

$U(\boldsymbol{R})$ is the projected potential of the column,

$$E = \frac{\hbar^2 k^2}{2m} \tag{82}$$

the incident electron energy, and λ is the electron wavelength. For $E_n < 0$ the eigenstates are bound to the column. We now rewrite Eq. (79) as

$$\psi(\boldsymbol{R}, z) = \sum_n C_n \phi_n(\boldsymbol{R}) + \sum_n C_n \phi_n(\boldsymbol{R}) \times \left[\exp\left(-i\pi \frac{E_n}{E} \frac{z}{\lambda}\right) - 1\right] \tag{83}$$

The coefficients C_n are determined from the boundary condition

$$\sum_n C_n \phi_n(\boldsymbol{R}) = \psi(\boldsymbol{R}, 0) \tag{84}$$

In the case of plane wave incidence one thus has

$$\sum_n C_n \phi_n(\boldsymbol{R}) = 1 \tag{85}$$

so that

$$\psi(\boldsymbol{R}, z) = 1 + \sum_n C_n \phi_n(\boldsymbol{R}) \times \left[\exp\left(-i\pi \frac{E_n}{E} \frac{z}{\lambda}\right) - 1\right] \tag{86}$$

Only states will appear in the summation, for which

$$|E_n| \geqslant \frac{E\lambda}{z} \tag{87}$$

These are bound states with deep energy levels, which are localized near the column cores. In practice if the atom column is not too heavy and the distance between columns not too close (e.g., larger than 0.1 nm) only one eigenstate will appear, which can be compared to the 1s state of an atom.

We then have

$$\psi(\boldsymbol{R}, z) = 1 + C\phi(\boldsymbol{R}) \times \left[\exp\left(-i\pi \frac{E}{E_0}\frac{z}{\lambda}\right) - 1\right] \tag{88}$$

A very interesting consequence of this description is that, since the state ϕ is very localized at the atom core, the wavefunction for the total object can be expressed as a superposition of the individual column functions ϕ_i so that Eq. (88) in that case becomes

$$\psi(\boldsymbol{R}, z) = 1 + \sum_i C_i \phi_i(\boldsymbol{R} - \boldsymbol{R}_i) \times \left[\exp\left(-i\pi \frac{E}{E_0}\frac{z}{\lambda}\right) - 1\right] \tag{89}$$

where the summation runs over all the atomic columns of the object, parallel to the electron beam.

The interpretation of Eq. (89) is simple. Each column i acts as a channel in which the wavefunction oscillates periodically with depth. The periodicity is related to the 'weight' of the column, that is, proportional to the atomic number of the atoms in the column and inversely proportional to their distance along the column. The importance of these results lies in the fact that they describe the dynamic diffraction for larger thicknesses than the usual phase grating approximation and that they require only the knowledge of one function ϕ_i per column (which can be tabulated similar to atom scattering factors or potentials). Furthermore, even in the presence of dynamical scattering, the wavefunction at the exit face still retains a one-to-one relation with the configuration of columns for perfect crystals as well as for defective crystals as far as they consist of columns parallel to the electron beam. Hence this description is very useful for interpreting high resolution images. Equation (89) is valid for light columns, such as Si(1 1 1) or Cu(1 0 0) with an accelerating voltage up to about 300 keV. When the atom columns are 'heavier' and/or the accelerating voltage higher or when the atom columns are very close (0.1 nm), the basic concept of channeling remains valid but the expression (89) has to be modified.

Figure 50 shows the electron density $|\psi(\boldsymbol{R}, t)|^2$ as a function of depth in an Au_4Mn alloy crystal for 200 keV incident electrons. The corners represent the projection of the manganese column. The square in the center represents the four gold columns. The distance between adjacent columns is 0.2 nm. The periodicity along the direction of the column is 0.4 nm. From these results it is clear that the electron density in each column fluctuates nearly periodically with depth. For gold this periodicity is about 4 nm, and for manganese 13 nm. These periodicities are nearly the same as for isolated columns so that the influence of neighboring columns in this case is still small. The energies of the respective s states are respectively about 250 and 80 eV.

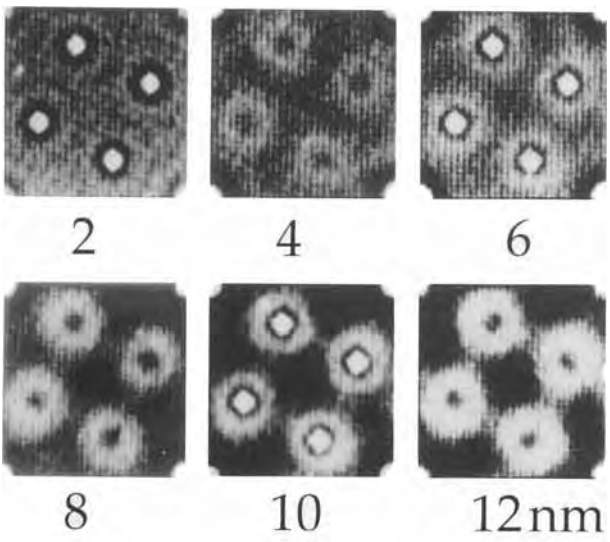


Figure 50. Electron density as a function of depth in Au_4Mn (see text).

It has to be stressed that the derived results are only valid in a perfect zone axis orientation. A slight tilt can destroy the symmetry and excite other, non-symmetric states, so that the results become much more complicated. It is interesting to note that channeling has usually been described in terms of Bloch waves. However, as follows from the foregoing, channeling is not a mere consequence of the periodicity of the crystal but occurs even in an isolated column parallel to the beam direction.

In this view, the Bloch wave approach is only of mathematical importance.

Diffraction Pattern

Fourier transforming the wavefunction (89) at the exit face of the object yields the wavefunction in the diffraction plane, which can be written as

$$\psi(\boldsymbol{g}, z) = \delta(\boldsymbol{g}) + \sum_i \exp(-2\pi i \boldsymbol{g} \cdot \boldsymbol{R}_i) F_i(\boldsymbol{g}, t) \tag{90}$$

In a sense the simple kinematical expression for the diffraction amplitude holds, provided the scattering factor for the atoms is replaced by a dynamical scattering factor for the columns, in a sense as obtained in [42] and which is defined by

$$F_i(\boldsymbol{g}, z) = \left[\exp\left(\frac{-i\pi E_i}{E}\,\frac{z}{\lambda}\right) - 1\right] C_i f_i(\boldsymbol{g}) \tag{91}$$

with $f_i(\boldsymbol{g})$ the Fourier transform of $\phi_i(\boldsymbol{R})$. It is clear that the dynamical scattering factor varies periodically with depth. This periodicity may be different for different columns.

In the case of a monoatomic crystal, all F_i are identical. Hence $\psi(\boldsymbol{g}, z)$ varies perfectly periodically with depth. In a sense the electrons are periodically transferred from the central beam to the diffracted beams and back. The periodicity of this dynamic oscillation (which can be compared with the Pendelösung effect) is called the dynamic extinction distance. It has for instance been observed in Si(1 1 1). An

important consequence of Eq. (90) is the fact that the diffraction pattern can still be described by a kinematical type of expression so that existing results and techniques (e.g., extinction rules) that have been based on the kinematical theory remain valid to some extent for thicker crystals in zone orientation.

Parameterization

Suppose all atom potentials would have the same form multiplied with a constant proportional to the atomic number Z one could expect that the eigenstate $\phi(\boldsymbol{r})$ and its energy E could scale with Z/d with d the repeat distance of atoms in the column. In this way, the wavefunction could be completely parameterized in terms of Z/d and Z. Even if Eq. (88) is only qualitatively correct, this would allow the position of each column to be determined very accurately, and the 'weight' of the column and the crystal thickness approximately. In practice it has been found that the bound state scales with $Z^{1/2}$ for all atoms of the periodic table (calculated using Smith and Burge parameters) that E roughly scales with Z/d. In this way, a robust parameterization is possible which enables the retrieval of the projected structure from matching with the exit wavefunction [43].

Appendix E: Phase Retrieval

Consider an image plane at a particular focus value for which we want to retrieve the phase. For convenience we choose the origin of focus in that plane. Writing the wavefunction as a Fourier integral, we have

$$\psi(\boldsymbol{R},0) = C + \int_{\boldsymbol{g}\neq 0} \phi(\boldsymbol{g}) \exp(2\pi i \boldsymbol{g}\cdot\boldsymbol{R})\, \mathrm{d}\boldsymbol{g} \tag{92}$$

where $\phi(\boldsymbol{g})$ are the Fourier components. We have separated the zeroth-order component (constant term).

The wavefunction at a particular focus value z can be obtained from Eq. (92) by propagation, that is,

$$\psi(\boldsymbol{R},z) = C + \int_{\boldsymbol{g}\neq 0} \phi(\boldsymbol{g}) \exp(2\pi i \boldsymbol{g}\cdot\boldsymbol{R}) \times \exp(-i\pi\lambda g^2 z)\, \mathrm{d}\boldsymbol{g} \tag{93}$$

Three-dimensional Fourier transforming the image intensity $|\psi(\boldsymbol{R},z)|^2$ now yields, using Eq. (93),

$$\begin{aligned} I(\boldsymbol{g},\xi) = {} & |C|^2\delta(\boldsymbol{g}) + C^*\phi(\boldsymbol{g})\delta\left(\xi - \frac{\lambda g^2}{2}\right) \\ & + C\phi^*(-\boldsymbol{g})\delta\left(\xi + \frac{\lambda g^2}{2}\right) \\ & \times \int_{\substack{\boldsymbol{g}'\neq 0 \\ \boldsymbol{g}+\boldsymbol{g}'\neq 0}} \mathrm{d}'\boldsymbol{g}\phi^*(\boldsymbol{g}')\phi(\boldsymbol{g}+\boldsymbol{g}') \\ & \times \delta\left(\xi - \frac{\lambda}{2}[(\boldsymbol{g}+\boldsymbol{g}')^2 = \boldsymbol{g}'^2]\right) \end{aligned} \tag{94}$$

where δ are Dirac functionals and $\boldsymbol{g}$ and ξ are the conjugates of $\boldsymbol{R}$ and z. The first term on the right-hand side of Eq. (94) only yields a contribution in the origin. The second and third linear terms give a sharply peaked contribution which is located on a paraboloid in reciprocal space which can be considered as the Ewald sphere in a vacuum (Fig. 18). The last term gives a contribution which is more continuously spread through reciprocal space. It is immediately clear that by selecting the information concentrated on the

paraboloid one directly obtains $\phi(\boldsymbol{g})$ and $\phi^*(\boldsymbol{g})$, so that from Eq. (92) the total wavefunction at focus 0 is retrieved. In principle, this can be done by taking a nearly continuous series of images at very small focus intervals, three-dimensional Fourier transforming, and selecting the sphere. However, this procedure is rather impractical. Hence we proceed as follows. We take a series of images at focus values $z_1, z_2, z_3, \ldots$ The focus interval is of the order of 10 nm. Each of the images $I(\boldsymbol{R}, z_n)$ is then transformed into $I(\boldsymbol{g}, z_n)$. Finally we calculate the series

$$\sum_n I(\boldsymbol{g}, z_n) \exp(-i\pi\lambda g^2 z_n)$$

In this way, the sphere for $\xi > 0$ is in a sense projected in the plane, apart from a known weighting factor. We can do the same for $\xi < 0$. From these data it is easy to calculate $\phi(\boldsymbol{g})$. In a sense, all the images are back propagated to zero focus, where the linear part of each image superimposes and increases with respect to the nonlinear part. However, as seen from Eq. (94), the integral also gives a contribution to the sphere which may influence the results. This contribution can be corrected for by using Eq. (94) in an iterative way. Another advantage of this method is that it is relatively easy to compensate for the effect of chromatic aberration. It is well known that chromatic aberration results from a spread in the focus due to instabilities in voltage and lens current. Hence the image intensity is convoluted with a focal spread function $I(\boldsymbol{R}, z) = I_0(\boldsymbol{R}, z) * f(z)$. In reciprocal space, the convolution product is a direct product with the Fourier transform of $f(z)I(\boldsymbol{g}, \xi) = I_0(\boldsymbol{g}, \xi)f(\xi)$. If the spread function f is known, it is easy to compensate for this effect by dividing $f(\xi)$. Since this has to be done only at the sphere, blowing up effects are largely reduced. Another advantage of the method is that, since the contribution of the noise is more homogeneously distributed in space, the selection of the sphere automatically increases the signal-to-noise ratio. In practice the whole procedure is performed in two steps. A first approximation of the exit wave is obtained by the paraboloid method described above; then this result is refined using a maximum likelihood method as proposed by Coene et al. [16].

Acknowledgments

The author wishes to thank J. Janssen, M. Op de Beeck, D. Tang, A. Thust, J. Van Landuyt and G. Van Tendeloo for the use of material and/or stimulating discussion.

1.1.2.9 References

[1] O. Scherzer, *J. Appl. Phys.* **1949,** *20*, 20; J. C. H. Spence, *Experimental High Resolution Electron Microscopy*, Clarendon Press, Oxford **1988.**

[2] D. Van Dyck, A. F. De Jong, *Ultramicroscopy* **1992,** *47*, 266; A. F. De Jong, D. Van Dyck, *Ultramicroscopy* **1993,** *49*, 66; D. Van Dyck, M. Op de Beeck, W. M. J. Coene, *MSA Bull.* **1994,** *24*(2), 427.

[3] H. Rose, *Proc. XIIth International Congress on Electron Microscopy* (*Seattle*), San Francisco Press, San Francisco **1990,** pp. 202–203; H. Rose, *Optik* **1990,** *85*, 19.

[4] W. O. Saxton, *J. Microsc.* **1995,** *179*(2), 210.

[5] O. Krivanek, *Ultramicroscopy* **1994,** *55*, 419.

[6] D. W. Robards, A. J. Wilson, *Procedures in Electron Microscopy*, Wiley, Chichester **1993.**

[7] J. M. Cowley, S. Iijima, *Z. Naturforsch.* **1972,** *27a*(3), 445.

[8] D. F. Lynch, A. F. Moodie, M. A. O'Keefe, *Acta Crystallogr.* **1975,** *A31*, 300.

[9] D. Van Dyck, G. Van Tendeloo, S. Amelinckx, *Ultramicroscopy* **1982,** *10*, 263; S. Amelinckx, G. Van Tendeloo, J. Van Landuyt, *Bull. Mater. Sci.* **1984,** *6*(3), 417; S. Amelinckx, G. Van Tendeloo, J. Van Landuyt, *Ultramicroscopy* **1985,** *18*, 395.

[10] C. Kisielowski, P. Schwander, F. H. Baumann, M. Seibt, Y. Kim, A. Ourmadz, *Ultramicroscopy* **1994,** *58*, 131.
[11] S. Amelinckx, *Acta Crystallogr.* **1995,** *B51*, 486.
[12] P. A. Stadelman, *Ultramicroscopy* **1987,** *21*, 131; Anon, *Multislice Computer Program*, High Resolution Electron Microscope Facility, Arizona State University, Tempe, AZ; R. Kilaas, R. Gronsky, *Ultramicroscopy* **1982,** *11*, 289; D. Van Dyck, W. Coene, *Ultramicroscopy* **1984,** *15*, 29; D. Van Dyck, W. Coene, *Ultramicroscopy* **1984,** *15*, 41; D. Van Dyck, W. Coene, *Ultramicroscopy* **1984,** *15*, 287.
[13] R. Bierwolf, M. Hohenstein, *Ultramicroscopy* **1994,** *56*, 32.
[14] J. Jansen, D. Tang, H. W. Zandbergen, M. Schenk, *Acta Cryst. A*, **1996,** in press.
[15] P. Schiske in *Image Processing of Computer-aided Design in Electron Optics* (Ed.: P. Hawkes) **1973**; W. O. Saxton in *Proceedings of the XIth International Congress on Electron Microscopy, Kyoto.*
[16] D. Van Dyck in *Proceedings of the XIIth International Congress on Electron Microscopy (Seattle)*, San Francisco Press, San Francisco **1990,** pp. 26–27; W. Coene, G. Janssen. M. Op de Beeck, D. Van Dyck, *Phys. Rev. Lett.* **1992,** *29*, 37; M. Op de Beeck, D. Van Dyck, W. Coene in *Electron Holography* (Eds.: A. Tonomura et al.), North-Holland/Elsevier, Amsterdam **1995,** pp. 307–316.
[17] M. A. O. Miedema, A. H. Buist, A. Vandenbos, *IEEE Trans. Instrum. Measur.* **1994,** *43*(2), 181.
[18] D. Tang, H. Zandbergen, J. Jansen, M. Op de Beeck, D. Van Dyck, *Ultramicroscopy* (in press).
[19] A. Thust, K. Urban, *Ultramicroscopy* **1992,** *45*, 23.
[20] D. Broddin, G. Van Tendeloo, S. Amelinckx, *J. Phys. C: Condens. Matter* **1990,** *2*, 3459.
[21] M. Takeda, G. Van Tendeloo, S. Amelinckx, *Acta Crystallogr.* **1988,** *A44*, 938.
[22] H. Alexander, J. C. H. Spence, D. Shindo, H. Gottschalk, N. Long, *Phil. Mag.* **1986,** *53*(5), 617.
[23] S. Kuypers, G. Van Tendeloo, S. Amelinckx, H. W. Shu, S. Faulmes, J. Flahaut, P. Laruelle, *J. Solid State Chem.* **1988,** *73*, 192.
[24] S. Kuypers, J. Van Landuyt, S. Amelinckx, *J. Solid State Chem.* **1990,** *86*, 212.
[25] R. Bontchev, B. Darrut, J. Parriet, F. Weill, G. Van Tendeloo, S. Amelinckx, *Eur. J. Solid State Inorg. Chem.* **1993,** *30*, 521.
[26] H. W. Zandbergen, W. A. Groen, F. C. Mijlhoff, G. Van Tendeloo, S. Amelinckx, *Physica* **1988,** *C156*, 325.
[27] T. Krekels, G. Van Tendeloo, S. Amelinckx, *Solid State Commun.* **1991,** *79*, 607.
[28] L. C. Nistor, G. Van Tendeloo, S. Amelinckx, *J. Solid State Chem.* **1994,** *109*, 152.
[29] S. Muto, G. Van Tendeloo, S. Amelinckx, *Phil. Mag.* **1993,** *B67*, 443.
[30] X. F. Zhang, X. B. Zhang, D. Bernaerts, G. Van Tendeloo, S. Amelinckx, J. Van Landuyt, H. Werner, *Ultramicroscopy* **1994,** *55*, 25.
[31] S. Amelinckx, D. Bernaerts, X. B. Zhang, G. Van Tendeloo, J. Van Landuyt, *Science* **1994,** *267*, 1334; S. Amelinckx, X. B. Zhang, D. Bernaerts, X. F. Zhang, V. Ivanov, J. B. Nagy, *Science* **1994,** *265*, 635.
[32] L. Nistor, G. Van Tendeloo, S. Amelinckx, C. Cros, *Phys. Status Solidi (a)* **1994,** *146*, 119.
[33] J. Frank, *Optik* **1973,** *38*, 519; P. L. Fejes, *Acta Crystallogr.* **1977,** *A33*, 109.
[34] M. Born, E. Wolf, *Principles of Optics*, Pergamon Press, London **1975,** Chap. X.
[35] D. Van Dyck, *Adv. Electron. Electron Phys.* **1985,** *65*, 295.
[36] J. M. Cowley, A. F. Moodie, *Acta Crystallogr.* **1957,** *10*, 609.
[37] K. Ishizuka, N. Uyeda, *Acta Crystallogr.* **1977,** *A33*, 740.
[38] D. Van Dyck, W. Coene, *Ultramicroscopy* **1984,** *15*, 29.
[39] J. Lindhard, *Mat. Fys. Medd. Dan. Vid. Selsk* **1965,** *34*, 1; A. Tamura, Y. K. Ohtsuki, *Phys. Status Solidi (b)* **1974,** *73*, 477; A. Tamura, F. Kawamura, *Phys. Status Solidi (b)* **1976,** *77*, 391; B. Buxton, J. E. Loveluck, J. W. Steeds, *Phil. Mag.* **1978,** *A3*, 259.
[40] K. Kambe, G. Lehmpfuhl, F. Fujimoto, *Z. Naturforsch.* **1974,** *29A*, 1034; D. Van Dyck, J. Danckaert, W. Coene, E. Selderslaghs, D. Broddin, J. Van Landuyt, S. Amelinckx in *Computer Simulation of Electron Microscope Diffraction and Images* (Eds.: M. Krakow, M. O'Keefe), TMS Publications, MMM Society **1989,** pp. 107–134; D. Van Dyck, M. Op de Beeck in *Electron Holography* (Eds.: A. Tonomura et al.), North-Holland, Amsterdam **1994,** pp. 297–307; M. V. Berry, K. E. Mount, *Rep. Progr. Phys.* **1972,** *35*, 315; B. F. Buxton, J. E. Loveluck, J. W. Steeds, *Phil. Mag.* **1978,** *38*(3), 259.
[41] K. Kambe, G. Lempfuhl, F. Fujimoto, *Z. Naturforsch.* **1974,** *29a*, 1034.
[42] D. Shindo, M. Hirabayashi, *Acta Crystallogr.* **1988,** *A44*, 954.
[43] D. Van Dyck, M. Op de Beeck, *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier Science B.V. **1995,** p. 297.
[44] S. Amelinckx, D. Van Dyck in *Electron Diffraction Techniques* (Ed.: J. M. Cowley), Vol. 2, *IUCr Monographs on Crystallography*, International Union of Crystallography, Oxford University Press, Oxford **1992,** pp. 1–222.
[45] H. Alexander, J. C. H. Spence, D. Shindo, H. Gottschalk, N. Long, *Phil. Mag.* **1986,** *A53*(5), 627.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

1.2 Reflection Electron Microscopy

1.2.1 Introduction

In the early days of electron microscopy, Ruska [1] attempted to obtain images by reflecting electrons from the surfaces of solids, with limited success. The idea was taken up by several groups in the 1950s [2, 3] and images of surfaces showing moderately good resolution were obtained. However, in order to get useful image intensity, it was necessary to use electrons scattered at small angles, 5 to 10 degrees, from the surface so that the images suffered from severe foreshortening. When SEI (secondary electron imaging) was developed and gave comparable resolution with normal beam incidence and no foreshortening, the reflection mode fell out of favor. The reason why it was later revived was that it was realized that, if the images were obtained using strong diffracted beams from crystal surfaces, the image contrast could be highly sensitive to small changes in crystal structure, orientation or composition of the top few atomic layers of the surface. Thus it could give information, not available from SEI, of great value in surface science.

The first images obtained in 100 keV electron microscopes using strong diffracted beams confirmed the sensitivity of the method to surface structure [4, 5]. Bulk crystals were mounted in the usual specimen stage of the microscope so that the electron beam was incident on a flat crystal surface at a grazing angle of a few degrees. The reflection high energy electron diffraction (RHEED) pattern could be observed in the selected area electron diffraction (SAED) mode and a strong diffraction spot (usually the specular beam) could be chosen to pass through the objective aperture and form the image. The reflection electron microscopy (REM) imaging mode could then be likened to that for the so-called 'high-resolution' mode of dark-field imaging in transmission. For a diffraction angle of θ, the incident beam is tilted by an angle of 2θ to the axis of the objective lens and the crystal face is tilted at an angle θ, so that the diffracted beam passes down the axis through a centered objective aperture.

Also it was realized that, as in the case of transmission microscopy, an equivalent scanning form of imaging (SREM) was possible. Initially, it was demonstrated that useful images could be obtained with medium-energy (5–15 keV) electrons in the scanning mode when the small electron probe from a field-emission gun was scanned over a surface and a strong diffracted beam was detected to form the image [6, 7].

For meaningful research in surface science it is necessary, for most materials,

to study surfaces prepared and maintained under ultrahigh vacuum (UHV) conditions. For imaging with moderate resolution and medium-energy electrons this may be achieved by adding a field-emission gun to an ultrahigh vacuum chamber with only a simple lens system to focus the electron probe on the specimen. Standard specimen manipulators could be used and the reflection imaging could be combined with an imaging of the surface using secondary or Auger electrons or analysis of the surface composition using AES or other methods. A number of useful systems have been developed along these lines and applied to surface structure problems [8–10]. However, the spatial resolution achievable in such systems is limited to 5–10 nm. To achieve normal electron microscope resolutions under clean surface conditions, it is necessary to build a UHV specimen environment into a TEM instrument. By doing this, and providing a variety of devices for in situ specimen treatment, Yagi and coworkers [11] showed that images of clean surfaces could be obtained with resolutions approaching 1 nm and other groups have done likewise [12]. Later it was demonstrated that a resolution of almost 0.3 nm could be achieved in a 1 MeV electron microscope [13].

The reflection electron microscopy images obtained in electron microscopes at 100 keV or more make use of diffracted beams making angles of only a few degrees with the surface and so are foreshortened by a factor of 20 to 50 times. The magnification and resolution of the images in the direction of the incident beam are then severely limited and the high resolution of the images is possible only in the direction across the surface, perpendicular to the beam. However, the value of the technique is immediately apparent, in spite of this disadvantage. Steps on a crystal surface only one atom high are clearly visible, as in Fig. 1. Strong contrast is shown between surface regions that differ only in the structure of the first one or two monolayers of atoms, as seen in Fig. 2. Small bumps, dips, impurities or projections on the surfaces become obvious (see Figs. 8 and 10). The changes in surface structure or composition resulting from in situ treatments can be followed in detail. Also the images may be correlated with RHEED patterns revealing the crystal structure and with microanalysis using EELS [14] or EDS to determine the chemical composition of surfaces with high spatial resolution.

REM images can be obtained using any TEM instrument if a bulk specimen is

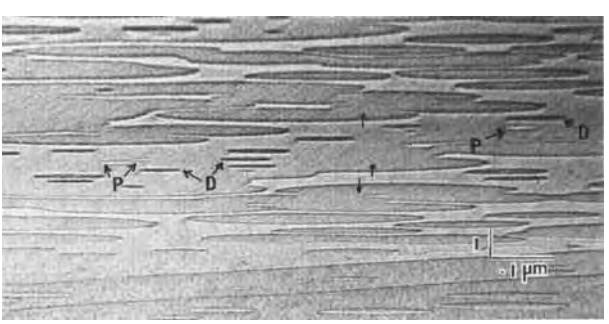


Figure 1. REM image of one-atom high surface steps on a Pt (1 1 1) surface. Arrows indicate the step-down direction. Circular features, foreshortened to thin ellipses, are atom-high projections, P, or depressions, D. Note the intensity changes across the steps. (Courtesy of T. Hsu.)

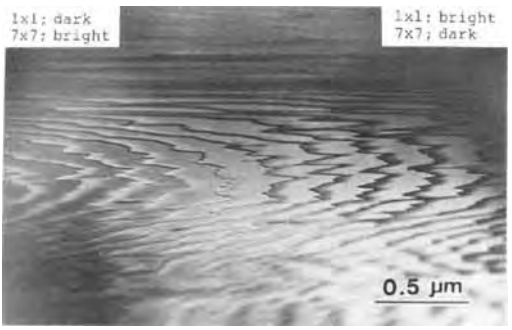


Figure 2. REM image of Si (1 1 1) face showing the formation of the 7 × 7 superstructure on the top side of surface steps. Differences in diffraction conditions give differences in contrast for the two sides of the image. (Reprinted with permission of K. Yagi [60].)

mounted so that the flat face of a crystal is almost parallel to the incident electron beam [15]. For many specimens such as noble metals and some refractory oxides the vacuum of about 10^{-7} torr of standard microscopes is not a serious hindrance to much research concerned with the structures of surfaces although, when chemical or physical surface reactions are to be studied, extreme precautions must be taken to avoid complications due to contamination of the surface. For semiconductors, most metals and other inorganic compounds, UHV is essential in most cases and access to one of the few microscopes modified for UHV operation is required. High voltage microscopes, operating at up to 1 MeV, can give excellent results [13] even though the foreshortening of the images is increased because of the smaller diffraction angles. Scanning transmission electron microscopy (STEM) instruments, which normally have a vacuum better than 10^{-8} torr, may also be used with the incident beam scanned over the surface at grazing incidence to give SREM images [16].

The main limitation on the resolution of REM images comes from the large energy spread of the diffracted electrons [17]. For the usual RHEED geometry, the average distance traveled in the crystal by an emerging 100 keV electron may be 100 nm or more; greater than the path-length for inelastic scattering processes such as the excitation of plasmons or the excitation of outer-shell or valence-band electrons of the solid, with energy losses of 10 to 30 eV. Also it may be noted that the electron beam traveling almost parallel to a specimen surface can excite surface plasmons while it is within 1 or 2 nm of the surface, on either side, and the energy losses from this source are in the range of 5 to 20 eV. The strong RHEED diffraction spots may contain 50 to 80% of electrons which have lost more than 3 eV of energy [18]. The average energy loss is of the order of 30 eV and the energy spread is about the same. The chromatic aberration of the objective lens in a 100 keV microscope then limits the resolution to about 1 nm. Better resolution can be obtained by use of higher voltage microscopes (up to 1 MeV) since the loss of resolution from this cause depends on the ratio of the energy loss to the initial electron energy. Otherwise, both the resolution and the contrast of REM images may be improved by use of an energy filter, such as is standard equipment for STEM instruments and is now available for some TEM instruments.

1.2.2 Reflection High Energy Electron Diffraction Patterns

The RHEED pattern produced when a high-energy electron beam strikes a crystal

surface at grazing incidence includes sharp strong diffraction spots lying on a set of concentric circles known as Laue circles, corresponding to the Laue zones in which the Ewald sphere cuts the set of parallel planes of reciprocal lattice points in reciprocal space. In practice, the strong spots do not appear sharp because they are normally grossly overexposed when recorded on photographic film and greatly broadened by halation. Also, with the overexposure, the diffuse background around the spots, arising from thermal diffuse scattering and from the small-angle energy-loss electron distribution, blackens the film and makes the spots seem large. In the background, there is a pattern of Kikuchi lines (sometimes referred to simply as K-lines) resulting from the diffraction of diffusely scattered electrons by the crystal lattice planes. The K-line patterns are accompanied in many cases by sets of parabolas and circles (Fig. 3).

The configuration of the sharp Bragg spots can be derived from a reciprocal space diagram such as Fig. 4a. Because the crystal is terminated at a surface, and most of the diffraction occurs in a thin surface layer, continuous lines of scattering power, perpendicular to the surface, pass through the sharp maxima around the reciprocal lattice spots. Hence a sharp spot is generated whenever the Ewald sphere cuts one of these lines, giving rise to sets of spots lying on the Laue circles, as suggested in Fig. 4b. The spots are very bright when the Ewald sphere cuts close to a reciprocal lattice point. The spot most commonly used for REM imaging is the specular reflection, given by the mirror image of the incident beam in the plane of the crystal surface. This spot is most intense when the incident beam is at the Bragg angle for lattice planes parallel to the surface. No spots appear below the 'shadow edge' which marks the intersection of the surface plane with the recording

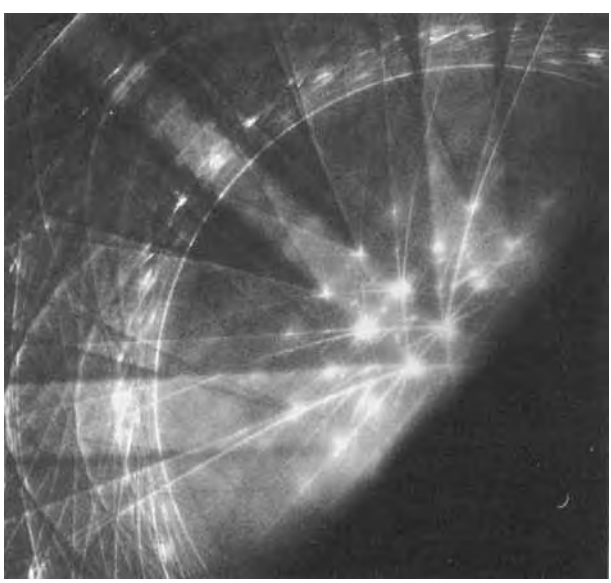


Figure 3. RHEED pattern from the (1 1 0) surface of GaAs showing Bragg spots, K-lines, parabolas and circles. (Reprinted with permission of K. Yagi [61].)

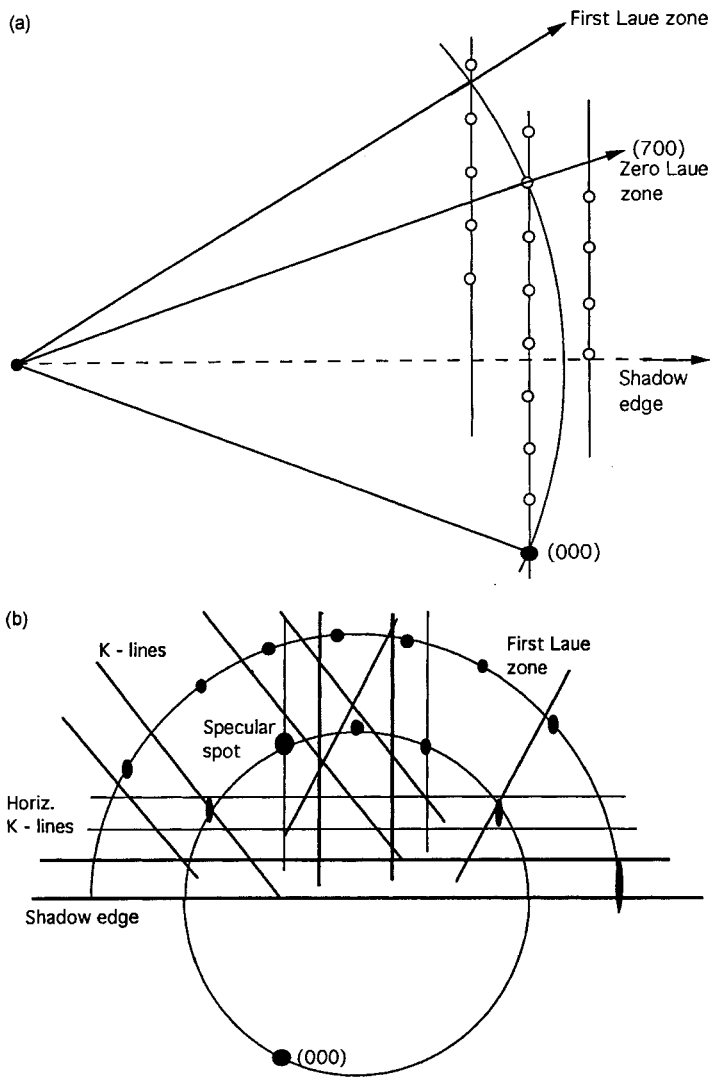


Figure 4. The geometry of RHEED patterns. (a) Section of reciprocal space in the plane of the incident and diffracted beams showing the intersection of the Ewald sphere with the reciprocal lattice. (b) The resulting pattern of spots and K-lines.

plate unless, as sometimes happens, there is transmission diffraction through small projections or edges of the crystal close to the termination of the crystal face.

Energy-loss processes involving the excitation of plasmons and single-electron excitations give rise to small-angle scattering (10^{-3} radians or less). The combination of such inelastic scattering with the elastic scattering gives the peaks of diffuse scattering around each Bragg reflection spot. The thermal diffuse scattering, which involves energy losses of the order of only 10^{-2} eV, gives high-angle scattering to 10^{-1} radians or more but is also peaked around the Bragg spots. Multiple scattering occurs for each process separately and between processes, so that the RHEED pattern has a broad diffuse background of, mainly, inelastically scattered electrons.

These electrons are diffracted by the crystal lattice, giving rise to the K-line patterns. These are analogous to the Kossel patterns produced when X-rays, generated inside a crystal, are diffracted by the lattice planes. However, in the case of high energy electrons, the wavelengths and diffraction angles are so small that, instead of the strongly curved Kossel lines, the K-lines produced appear as parallel sets of straight lines. The separation of a pair of parallel lines is twice the Bragg angle for diffraction from a set of crystal lattice planes, since a line is generated when the electrons are incident on the lattice planes at the Bragg angle from one side or the other. Multiple sets of parallel lines are generated, by several orders of diffraction from the same planes (see Figs. 3 and 4b). Usually a set of sharp horizontal K-lines is seen, parallel to the shadow-edge, corresponding to the diffraction from planes parallel to the crystal surface. The relative intensities and detailed intensity distributions of K-lines have been described in terms of dynamical diffraction theory for high-energy electrons by various authors [19].

Sets of K-lines, corresponding to sets of planes such as those associated with the reciprocal lattice points on a common reciprocal lattice plane, may be tangential to a common asymptotic circle or parabola. Strong parabolas and circles are seen in K-line patterns, close to, but slightly displaced towards the shadow edge from, these asymptotic curves. These parabolas and circles may be ascribed to a resonance phenomenon in which the diffusely scattered electrons are strongly confined to sets of planes or rows of atoms parallel to the crystal surface [20, 21]. When a convergent incident beam is used, sets of parabolas can also appear corresponding to channelling of the incident-beam electrons along planes perpendicular to the surface [22].

The resonance condition for planes parallel to the surface corresponds to the geometry for which a strong specular beam is produced because the incident beam is at the Bragg angle for planes parallel to the surface and also a strong diffracted beam is produced so that it runs almost parallel to the surface. This condition is indicated when the specular beam sits at the intersection of a horizontal K-line and either a strong inclined K-line, or, for diffusely scattered electrons, a strong parabola. Then the intensity of the specular reflection, and of the whole diffraction pattern, is greatly enhanced, so that this condition is often sought to produce high-intensity REM images. The resonance condition sometimes involves a channeling phenomenon in which the electron wave is channeled along the top one or two top surface layers of atoms and penetrates very little into the bulk of the crystal [23, 24]. Then the REM image becomes highly sensitive to the structure and composition of the top surface layers and the image contrast for surface defects is modified in ways which will be described later.

The geometry of the pattern of K-lines and circles and parabolas depends only on the orientation of the crystal relative to the recording film and is independent of the orientation of the incident beam since the pattern is derived from diffusely scattered electrons which may be considered as generated at points within the crystal. As a crystal face is rotated in azimuth, the K-line pattern rotates with it and hence gives a very sensitive indication of crystal

orientation. However, the incident beam direction determines the overall pattern of intensities of the scattered electrons and so governs the relative intensities and visibilities of the various lines and curves.

Because the refractive index of solids for high-energy electrons is slightly greater than unity, there is a refractive bending of all diffracted beams leaving a flat surface at a small angle, resulting in a displacement of all RHEED spots and lines toward the shadow edge. For marginal cases, the displacement may be as great as 10^{-2} radians but the displacement decreases rapidly for higher angles of diffraction. This introduces a distortion of the geometry of the RHEED pattern. The refraction effect also leads to a total internal reflection of some diffracted beams directed almost parallel to the surface, often contributing to the surface resonance effect.

1.2.3 Image Contrast for Surface Steps and Projections

For steps or projecting particles on an otherwise perfectly flat crystal face which are large compared with the resolution limit of the microscope, the REM image contrast can be described, to a good approximation, in terms of simple geometric optics. Fig. 5a, for example, illustrates the case for large steps for an exaggerated angle of incidence. For a down-step (as seen when looking towards the electron source) there is no discontinuity in the intensity diffracted from the crystal surface before the step and from the top of the step, so that the step is invisible apart from minor perturbations due to some transmission through the step edge. For an up-step, however, no electrons reach the image from the exit face of the

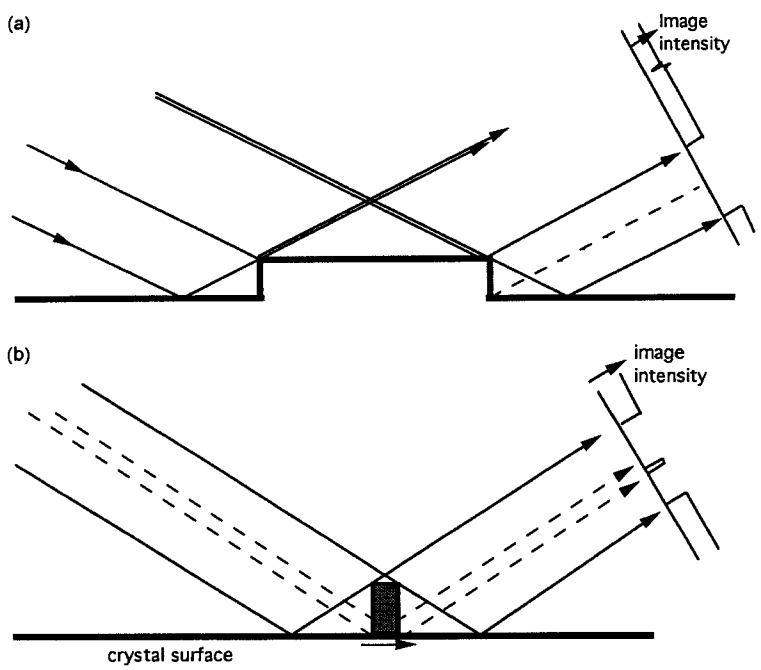


Figure 5. Geometric-optics diagrams of the diffracted beams from a large projection (a) and a thin projection or small particle, on a crystal surface and the resulting image intensities (b).

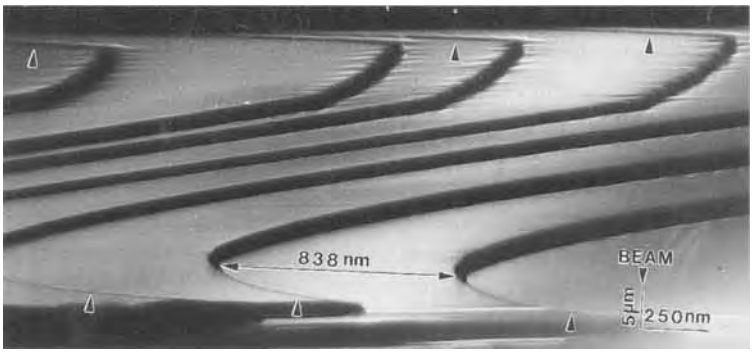


Figure 6. REM image of large curved steps and small intersecting steps (arrowed) on an oxygen-annealed rutile (1 0 0) surface [62].

step or from a region of the crystal surface after the step which is shaded from the incident beam. A black region therefore appears in the image of width equal to roughly twice the step height. The appearance of the step varies with the angle that the step makes with the incident beam direction. When the step becomes parallel to the incident beam direction, the width of the black shadow decreases to zero, and the form of the step is seen in profile. The consequent appearance of large curved steps in the image is as illustrated in Fig. 6.

For a projection from the surface which is thin in the beam direction, or for a small foreign particle sitting on the surface, the corresponding geometric-optics construction is as shown in Fig. 5b. The projection is illuminated by the electrons diffracted from the surface behind it (as seen when looking towards the electron source) and so forms an image as in transmission microscopy. The projection also prevents the incident beam from being diffracted from the surface in front of it. Hence the image of the projection is accompanied by an inverted mirror-image, as is evident in Fig. 7.

A common feature of such double images for particles on crystal surfaces is that the particle image and the mirror image are separated by a bright line. This effect has been attributed to a channeling phenomenon. Under the surface resonance conditions, electrons may be channeled to flow along the top one or two layers of atoms on the surface. Under this condition they may travel for unusually large distances along the surface before being diffracted out of the surface. The presence of a particle on the surface may actually enhance this effect since the presence of foreign atom layers on the surface may decrease the probability of the electrons escaping from the surface layers. Hence the electrons may pass underneath the

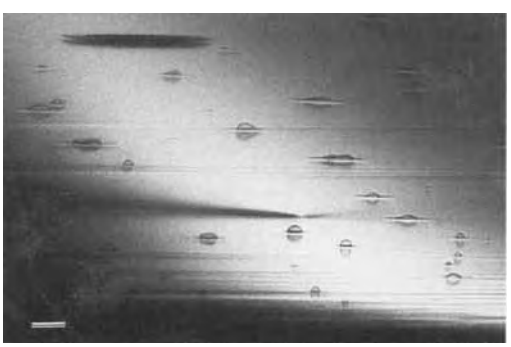


Figure 7. REM image of small transparent foreign particles on an Si (1 1 1) surface showing the direct image and mirror image and the bright line between them in each case (plus a dislocation strain field). Scale bar: approx. 50 nm. (Courtesy of T. Hsu.)

particle with little decrease of intensity and give a bright line in the image, as suggested in Fig. 5b.

For very low steps on a surface, the step height is often equal to the thickness of one or two monolayers of atoms, which is less than the resolution limit for most REM imaging. Then the geometric-optics pictures used for high steps can no longer apply and the strong contrast normally seen must be attributed to a phase-contrast mechanism. The two beams labeled 1 and 2 in Fig. 5a, coming from before the step and on top of the step, overlap and give rise to interference effects depending on their relative phase. For a step height h and with equal angles of incidence and reflection of the beams with the surface equal to $\theta_B + \Delta$, where θ_B is the Bragg angle (and ignoring refraction effects) the beams have a phase difference of $(2\pi/\lambda) \cdot 2\theta_B h(1 + \Delta/\theta_B)$ [25]. This is equal to $2\pi n(1 + \Delta/\theta_B)$ if the step height is n times the crystal plane spacing. If this phase difference is small, or close to a multiple of 2π, the contrast given by the step is the same as that at the edge of a weak phase object in a TEM image, that is, the contrast is zero for the objective lens exactly focused on the step (if the effect of the spherical aberration of the lens is ignored) but the out-of-focus images show a black–white contrast which reverses from overfocus to underfocus and gets stronger and broader as the amount of defocus increases. As for Fresnel fringes, the width of the dark and bright lines is proportional to the square root of the defocus.

Because of the glancing angle of incidence, the amount of defocus occurring in a normal REM image of an extended flat crystal face may be very large (many micrometers, positive and negative) so that strong contrast is seen over most of the field of view. If the step is an up-step rather than a down-step the same argument applies except that the sign of the phase difference is reversed and hence the dark-light contrast of the step is reversed. Hence an analysis of the image contrast for a through-focus series of images gives an immediate indication of whether the steps go up or down [11].

It is readily shown that if the incident beam is at the Bragg angle for the lattice planes parallel to the surface and the step height is equal to the lattice plane spacing, the phase difference of the beams scattered before and after the step is a multiple of 2π, giving no contrast. However, this can only occur if it can be assumed that the refraction effect is negligible, that the incident beam is exactly at the Bragg angle and that there is no variation of the lattice plane spacing at the step due to elastic strain effects. In most cases, if it is assumed that the potential function drops sharply at the crystal surface from the value in vacuum to the inner-potential value inside the crystal, the refraction effect at the surface ensures that the angle of incidence of the beam in vacuum is less than the Bragg angle for strong reflection within the crystal. Then the deviation, Δ, from the Bragg angle in the above equation may be quite large. Also, for many materials, it is thought that there is a relaxation effect around a step so that the step height deviates from the bulk interplanar spacing by a few percent. In either of these cases, or if the image is obtained at an incident angle which is not exactly the Bragg angle, the step is no longer equivalent to a weak phase object. A second order approximation then suggests that a sharp black line

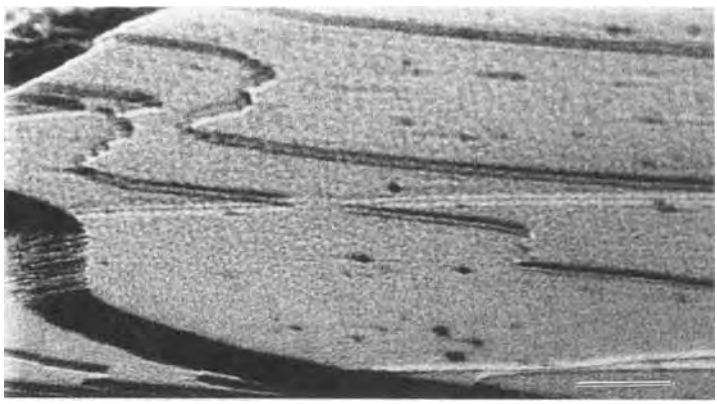


Figure 8. REM image of a Pt (1 1 1) surface, plus some small evaporated Au particles, showing the doubling of the images of small steps [63]. Scale bar: 40 nm.

may appear at the in-focus position, and for the images of the step with moderate amounts of defocus the antisymmetric dark-light contrast may be strongly perturbed [25].

In some cases the in-focus images of small steps appear as pairs of dark or light lines with a separation, in the image plane, of about 3 nm (Fig. 8). The step image doubling persists for out-of-focus images until it is swamped by the usual dark-light defocus contrast. This effect has been variously attributed to dynamical diffraction effects [26] or to strain fields due to lattice relaxation around the step [27]. However, a systematic investigation has shown that the effect is associated with the occurrence of surface resonance [28]. The doubling has been simulated by Anstis [29] using many-beam dynamical diffraction calculations.

1.2.4 Dislocations and Stacking Faults

Because the contrast in an REM image derives from the variation of intensity of a diffracted beam, the small changes of lattice orientation associated with local strain fields may be clearly visible. When a dislocation line emerges from the bulk and intersects a crystal surface, the associated strain fields distort the surface and so give rise to characteristic patterns of dark and light lobes which, when foreshortened, appear as streaks, extending across the surface, nearly perpendicular to the incident beam [30]. The sensitivity to lattice strain is such that the streaks may extend for distances of several micrometers (Fig. 9).

The nature of the dislocation may be derived from the configuration of the streaks and their variation with the diffraction conditions, in much the same way as in TEM [31]. For a pure screw dislocation normal to the surface, the surface distortion is similar to that of the lattice planes in bulk. When viewed from the incident beam direction, the surface lattice plains are tilted up on one side and down on the other side of the dislocation core. If the bulk crystal is exactly at the Bragg angle, a tilt in either direction decreases the intensity and gives a dark streak on each side. If the incident beam is slightly off the Bragg angle, the intensity is increased on one

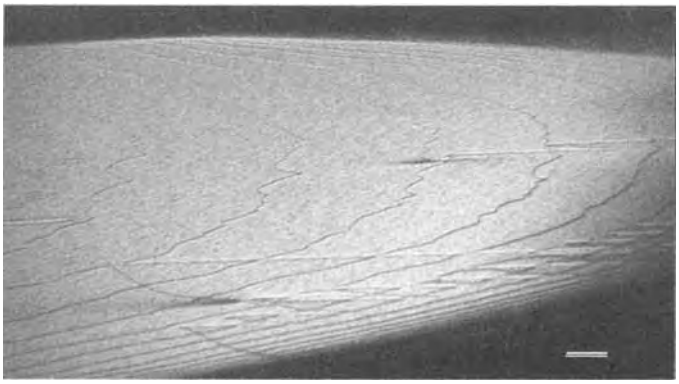


Figure 9. REM image of a Pt (1 1 1) surface showing slip trace steps from moving dislocations interacting with growth steps. Contrast from the dislocation strain fields appears at the terminations of the slip traces. Scale bar: approx. 50 nm. (Courtesy of T. Hsu.)

side and decreased on the other so that the dislocation gives an antisymmetrical, black–white pair of streaks.

An emerging edge dislocation gives a more complicated pattern of streaks. The well-known rule that the contrast vanishes for $\boldsymbol{g} \cdot \boldsymbol{b} = 0$, where $\boldsymbol{g}$ is the diffraction vector and $\boldsymbol{b}$ is the Burgers vector, holds for two-beam conditions, as in TEM, so that no contrast should be given for an edge dislocation perpendicular to the surface when imaged with a specular reflection. Even for this case, contrast may appear because the strain field varies the azimuthal rotation of the lattice relative to the beam and an azimuthal rotation can strongly influence the amplitudes of non-specular beams which may, in turn, affect the specular beam intensities through many-beam dynamical diffraction effects. Shuman [30] showed that edge dislocations can give images with four dark and light streaks.

When dislocations run parallel to a surface but below it, as in the case of dislocation loops which are wholly or partially buried, their strain fields may extend to the surface and perturb the orientations of the surface layers. Then diffuse, curved lines are seen in the REM image [32].

Frequently the contrast pattern indicating the emergence of a dislocation is seen at the termination of a step line. The step may be a growth step terminating in a screw dislocation or it may be a slip-trace, marking the passage of the dislocation as it has migrated through the crystal or glided along a fault plane (Fig. 9).

Planar stacking faults in the bulk crystal intersect the surface in straight lines which may or may not involve a surface step, depending on whether or not the displacement vector, $\boldsymbol{R}$, for the fault has a component perpendicular to the surface. In either case, the lattice displacement across the fault gives dark or light contrast, depending on the diffraction vector. Shuman [30] predicted that for a fault intercept with the surface, running perpendicular to the incident beam, there would be an oscillatory contrast similar to that seen in TEM images of faults but with contrast decreasing rapidly with depth of the fault below the crystal surface. However, no clear evidence of such contrast has been seen. If there is a surface step, the phase-contrast image, as described above, is added to the fault contrast and can be the predominant effect. For f.c.c. metals, fault lines have been seen joining the characteristic

contrast features corresponding to the two partial dislocations [33].

When f.c.c. metals such as Au and Pt are crystallized by rapid cooling from high temperatures, deformation fault lines or slip traces often appear on the surfaces, cutting across the curved growth steps as in Fig. 9. The points of intersection of the growth and fault steps are highly unstable and, if the sample has been at sufficiently high temperatures so that surface diffusion is possible, the sharp intersection points are rounded off, giving characteristic rounded cusps, lying along the straight line of the fault step, as seen in the image [34].

1.2.5 Surface Layers, Superlattices and Lattice Fringes

The high sensitivity of REM contrast to surface structure was first illustrated by the images obtained by Osakabe et al. [35], showing the growth of regions of the high-temperature Si (1 1 1) 7 × 7 structure on an Si (1 1 1), 1 × 1 face. The 7 × 7 structure appeared as dark or light bands along the low sides of growth steps (see Fig. 2), showing, for the first time, the importance of the steps as growth sites for this surface superlattice. Even more striking contrast was given when gold was evaporated on the Si (1 1 1) surface, giving an Au 5 × 1 superlattice structure with one third of a monolayer of Au. The Au superlattice was formed in domains in each of the three possible equivalent orientations and the different domain orientations gave very different image intensities [36]. With medium-energy SREM imaging, a similar contrast for different domain orientations was shown for Si (1 1 1) 2 × 1 superlattices [37].

Since then many such observations of surface structures and their domain configurations have been made. The notable advantage of the REM–RHEED method, as compared with the diffraction techniques such as LEED or other imaging methods such as LEEM or SEM is that the domain structures and their interactions with surface steps or other surface features may be examined with a spatial resolution of 1 nm or better.

For some oxide crystal faces, such as the Al_2O_3 (0 1 1) surface imaged in Fig. 10, areas of very different intensity are separated by steps, whereas other, darker steps separate regions of the same intensity [38]. It is believed that these observations arise because the surface plane may cut the unit cell of the crystal at various levels giving different terminations of the lattice. The differences in contrast across the steps depend on whether the height of the steps is equal to a fraction, or the whole, of the unit cell repeat distance. In the case of Fig. 10, the darker areas appear rough, whereas the lighter areas are smooth. It is suggested that the difference of intensity is due not just to the difference in lattice termination, but to a difference in chemical reactivity of the exposed surface atom layers which leads to a reaction with the remanent gases in the relatively poor vacuum of a normal TEM for some terminations.

The case is somewhat different for the (1 1 1) faces of noble metals where, again, there are intensity differences across surface steps [39] (see Fig. 1). Here surfaces terminated at different levels within a unit

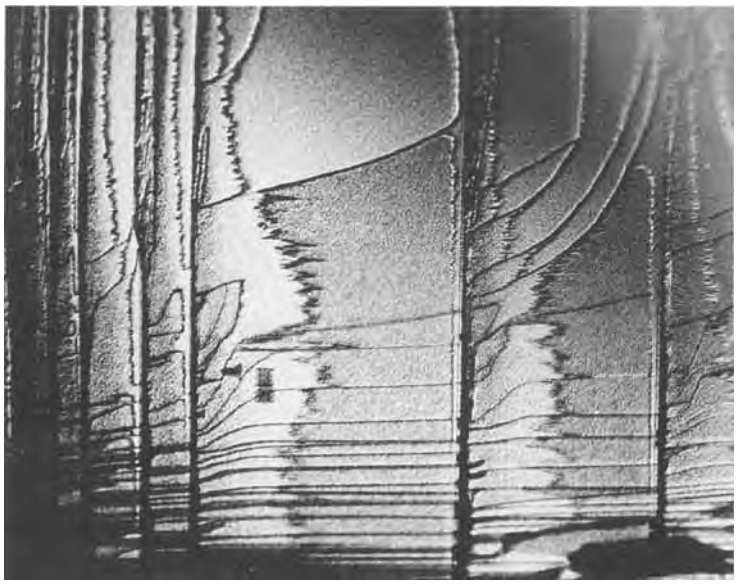


Figure 10. REM image from the (0 1 1) surface of an α-alumina crystal showing regions with differences of intensity corresponding to different surface terminations and step images which are different for step-heights of whole or partial unit cells [38].

cell are exactly equivalent. The contrast is explained if there are stacking faults parallel to the surface in the near-surface planes; for example, if the top three planes have an h.c.p. sequence rather than an f.c.c. sequence. Then, when an atom-high step occurs, there can be a change from one sequence to the other and hence a change of diffraction intensities.

Many cases of surface reconstruction involve the formation of surface superlattices. When the superlattice periodicities exceed the resolution limit for the REM imaging, lattice fringes of the corresponding periodicity become visible. For the Si (1 1 1) 7×7 structure the fringes corresponding to the 2.3 nm periodicity are clearly seen in such images as Fig. 11, due to Tanishiro et al. [40]. In later work with the high voltage UHV microscope, the basic lattice spacing of 0.34 nm was resolved by Koike et al. [13]. In images like these it is possible to observe the interactions of the periodic structures with the

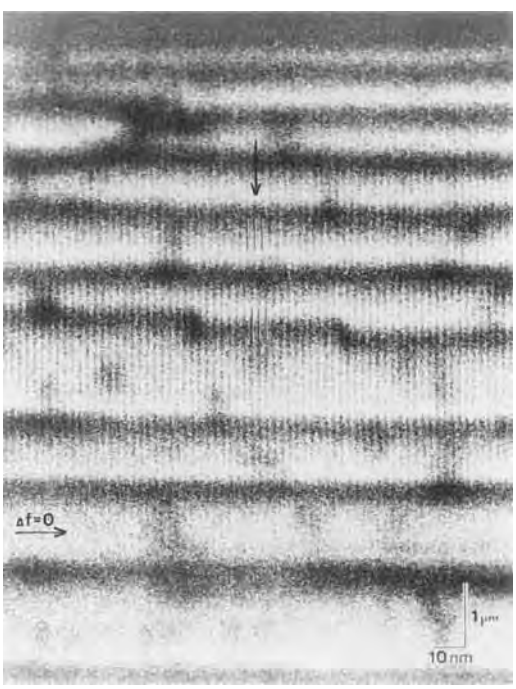


Figure 11. Lattice fringes of the 7×7 structure on an Si (1 1 1) face crossing surface steps which may, or may not, be out-of-phase boundaries of the 7×7 structure. (Courtesy of K. Yagi [40].)

surface defects such as steps, emerging dislocations and impurities. Superlattice fringes with periodicities in the range of 1–5 nm have also been observed for metals by Lehmpfuhl and Uchida [41] and for various oxides by Liu et al. [42].

The observation of surface lattice fringes in REM, however, is not quite as straightforward as for the TEM case. The fringes of Fig. 11 are obtained when the objective aperature transmits a 1×1 reflection from the underlying crystal plus several closely-spaced spots due to the surface superlattice, lying in a row parallel to the shadow-edge of the RHEED pattern. The superlattice spots lie on the Laue circle and so are often not in a line parallel to the shadow edge. If a group of superlattice spots which are on a line inclined to the shadow-edge are used to form the image, complications arise. The foreshortening factor and therefore the magnification in the beam direction, is different for each spot.

1.2.6 RHEED and REM Theory: Quantitative Interpretations

To allow quantitative interpretations of RHEED or REM intensities, two conditions must be met. Experimentally, the diffraction conditions must be determined with high precision. Then calculations must be made using an adequate many-beam dynamical diffraction formulation for the likely models of the structure. The dynamical diffraction theory is more complicated for the reflection case than for transmission of high-energy electrons, but a number of adequate approaches are now available.

The original dynamical theory of electron diffraction was formulated by Bethe [43] for the LEED case and was adapted in its simplest two-beam form by Collela [44] for RHEED and by Shuman [30] to describe the contrast given by crystal defects in REM images. However, the Bethe theory assumes that a perfect crystal lattice is cut off discontinuously at a planar boundary. It must be considerably modified to include the gradual decay of the crystal potential into the vacuum and the perturbation of the surface structure by relaxation, reconstructions or chemically modified layers. Also a many-beam form is usually essential.

Current theoretical treatments may be divided into two types. In one, the crystal potential distribution is divided into thin layers, parallel to the surface. Then the propagation of the electron wave into and out of the crystal is considered in terms of its modification by each layer and its transfer between layers [45–47]. This approach works very well for crystals which are perfect in the two dimensions of the flat surface, with relatively small periodicities. As the surface unit cell dimensions become larger, the number of beams involved increases rapidly and the size of the computation grows even more rapidly.

For nonperiodic objects, such as surface steps or boundaries of surface reconstructions, it is necessary to assume an artificial periodicity, using the assumption of periodic continuation. The assumption of large periodicity in real space corresponds to the sampling at small intervals in reciprocal space of the diffuse scattering arising from the nonperiodic object. The

repeat distance along the surface which must be considered is very large because of the small angles of incidence in REM, so that the number of sampling points in reciprocal space, and so the effective number of 'beams' in the calculation, is very large. However, even for this case, some successful calculations have been made by McCoy and Maksym [48].

In the other type of treatment, thin slices of the crystal perpendicular to the surface and almost normal to the incident beam are considered and transmission through these slices is simulated using the same sort of multislice, forward-scattering theory, as introduced by Goodman and Moodie [49], and used very successfully for transmission through thin crystals. The effect of the potential distribution in each slice of the specimen is given by a simple phase-object approximation since for forward scattering a small-angle approximation may be made. Then the propagation of the electron wave between slices is described in terms of Fresnel diffraction theory by convolution with a propagation function. The structure within each slice of the crystal in this case, however, is essentially non-periodic, since the slice cuts across the surface, and so an artificial periodicity must be assumed, as if the diffraction were from the faces of a set of parallel crystal slabs [50, 51]. A large number of beams must be considered even for a sharply terminated perfect crystal, but then only a small further complication is introduced by modifications of the surface structure. Also it is easy to introduce variations of structure in the incident beam direction, such as surface steps or projections or crystal faults, by varying the content of successive slices [24] (Fig. 12). This approach has been further modified and extended and applied to the simulation of REM images of surface steps and faults by Ma and Marks [52] and Anstis [29].

The fact that RHEED spots and REM images normally contain a high proportion of inelastically scattered electrons

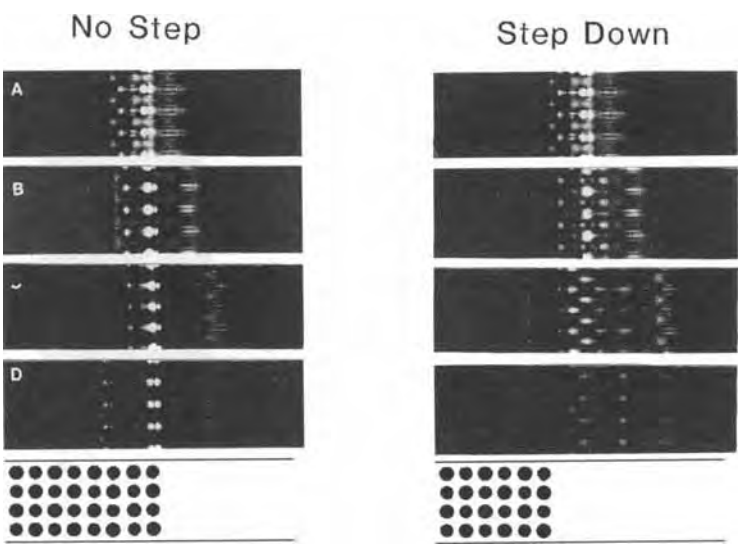


Figure 12. Calculations of the electron intensity distributions in planes perpendicular to the (1 0 0) surface of an MgO crystal at a distance, along the crystal face, of (A) 180, (B) 240, (C) 300 and (D) 480 slices, for a slice thickness of 0.2105 nm, after the entry of a narrow 100 kV electron beam into the crystal to give the (4 0 0) specular reflection. For the right-hand figures, there is a step-down after slice number 220. The bottom figures show the atom positions. The channeling of electrons along the surface planes of atoms is evident. (Courtesy of Z. L. Wang.)

introduces a complication which can be treated theoretically by including the inelastic scattering processes in the simulations; a difficult process [53]. It can be treated experimentally by use of an energy filter to remove the inelastically scattered electrons, implying that the calculations can be carried out for elastic scattering only but modified by the use of appropriate absorption functions.

1.2.7 Applications of Reflection Electron Microscopy

Of the many successful applications of REM imaging, a few may be mentioned as illustrating the scope and potential of the method. The capability for imaging surface steps has led to studies of step movements in the processes of crystal growth and evaporation, and of the movements of steps under the influence of applied electric fields by Yagi et al. [54]. When a current flows through a specimen of Si, the surface steps may be seen to bunch together or spread apart, depending on the direction of the current, indicating that there is an essential asymmetry in the surface structure. Likewise, the migration of metal atoms on the Si surfaces may depend on the current direction. Such surprising results have important inferences for the understanding of surface migration processes.

The observation of steps and their decoration has provided evidence regarding the importance of steps for crystal growth processes. Steps frequently form the nuclei for the condensation of impurity atoms arriving from the ambient atmospheres or by migration from the bulk of the specimen, as seen by Crozier et al. [55]. Small crystals growing on crystal surfaces as a result of deposition or chemical reaction are frequently seen to nucleate at steps or other defects of the surface [12].

The morphology of surfaces, described in terms of roughness or smoothness on an atomic scale, the occurrence of large steps and facets, or the frequency and regularity of small steps (Figs. 6, 8, and 10), provides information of importance in relation to the use of the surfaces as bases for epitaxial growth of thin films, for example of magnetic films or superconductors, for technical purposes. The REM technique has the advantage in such assessments, of allowing rapid surveys of large areas, plus studies of selected areas with high spatial resolution. Studies have been made by Wang and Bentley [56] at quite high temperatures, in situ, to show the changes of surface morphology taking place during the process of annealing of ceramic surfaces.

It has recently been shown that domain boundaries intersecting the surfaces of ferroelectric crystals can be seen and their movements under applied electric fields can be observed [57]. In this way it may be possible to throw light on a number of questions relating to the relevance of surface domain structures in relation to ferroelectric device properties.

A development of great potential significance has been the application of electron holography in the REM mode by Osakabe et al. [58] and Banzhof et al. [59]. As in transmission off-axis holography (Sec. 1.8 of this Chapter), an electrostatic biprism is inserted near to the image plane of the objective lens of a TEM

instrument. The electron waves reflected from the area of interest, the region of a flat crystal surface containing a defect, are made to interfere with the reference wave which is the wave reflected from an adjacent area of perfect crystal face. Perturbations of the pattern of interference fringes indicate the differences in phase of the two waves due to differences of height of the local surface relative to the ideally perfect crystal surface. Hence it is possible to map the local distortions of the surface. It has been shown that, using this technique, measurements of step heights or the strain fields around emerging dislocations may be made with an accuracy of 0.01 nm or better.

1.2.8 References

[1] E. Ruska, *Z. Phys.* **1933,** *83*, 492.
[2] J. W. Menter, *J. Photogr. Sci.* **1953,** *1*, 12.
[3] K. Ito, T. Ito, M. Watanabe, *J. Electron. Microsc.* **1954,** *2*, 10.
[4] J. S. Halliday, R. C. Newman, *Br. J. Appl. Phys.* **1960,** *11*, 158.
[5] P. E. H. Nielsen, J. M. Cowley, *Surf. Sci.* **1976,** *54*, 340.
[6] J. M. Cowley, J. L. Albain, G. G. Hembree, P. E. H. Nielsen, F. A. Koch, J. D. Landry, H. Shuman, *Rev. Sci. Instrum.* **1975,** *46*, 826.
[7] E. S. Elibol, H.-J. Ou, G. G. Hembree, J. M. Cowley, *Rev. Sci. Instrum.* **1985,** *56*, 1215.
[8] T. Ichinokawa, *Ultramicroscopy* **1984,** *15*, 193.
[9] M. Ichikawa, T. Doi, M. Ichihashi, K. Hayakawa, *Jpn. J. Appl. Phys.* **1984,** *23*, 913.
[10] P. A. Bennett, A. P. Johnson in *Reflection High Energy Electron Diffraction and Reflection Imaging of Surfaces* (Eds.: P. K. Larsen, P. T. Dobson), Plenum Press, New York **1988,** p. 371.
[11] N. Osakabe, Y. Tanishiro, K. Yagi, G. Honjo, *Surf. Sci.* **1981,** *102*, 424.
[12] A. V. Latyshev, A. B. Krasilnikov, A. L. Aseev, *Micros. Res. Tech.* **1992,** *20*, 341.
[13] H. Koike, K. Kobayashi, S. Ozawa, K. Yagi, *Jpn. J. Appl. Phys.*, **1989,** *28*, 861.
[14] Z. L. Wang, J. Bentley, *Microsc. Microanal. Microstruct.* **1991,** *2*, 301.
[15] T. Hsu, *Microsc. Res. Tech.* **1992,** *20*, 318.
[16] J. Liu, J. M. Cowley, *Ultramicroscopy* **1993,** *48*, 381.
[17] J. M. Cowley, *Ultramicroscopy* **1992,** *47*, 187.
[18] Z. L. Wang, J. M. Cowley, *Surf. Sci.* **1988,** *193*, 501.
[19] K. Okamoto, T. Ichinokawa, Y.-H. Ohtsuki, *J. Phys. Soc. Jpn.* **1971,** *30*, 1690.
[20] A. G. Emslie, *Phys. Rev.* **1934,** *45*, 43.
[21] M. Gajdardziska-Josifovska, J. M. Cowley, *Acta Crystallogr.* **1991,** *A47*, 74.
[22] L. Wang, J. M. Cowley, *Ultramicroscopy* **1994,** *55*, 228.
[23] H. Martin, G. Meyer-Ehmsen, *Surf. Sci.* **1985,** *151*, 570.
[24] Z. L. Wang, P. Lu, J. M. Cowley, *Ultramicroscopy* **1987,** *23*, 205.
[25] J. M. Cowley, L.-M. Peng, *Ultramicroscopy* **1985,** *16*, 59.
[26] Y. Uchida, G. Lehmpfuhl, *Ultramicroscopy* **1987,** *23*, 53.
[27] L.-M. Peng, J. M. Cowley, T. Hsu in *Proceedings of the 45th Annual Meeting of the Electron Microscopy Society of America* (Ed.: G. W. Bailey), San Francisco Press, San Francisco **1987,** p. 34.
[28] N. Yao, J. M. Cowley, *Microsc. Res. Tech.* **1992,** *20*, 413.
[29] G. R. Anstis, in *Proc. Internat. Conf. Electron Micros.-13*, Paris **1994,** p. 1027.
[30] H. Shuman, *Ultramicroscopy* **1977,** *2*, 361.
[31] L.-M. Peng, J. M. Cowley, *Micron Microscopia Acta,* **1987,** *18*, 171.
[32] L.-M. Peng, J. M. Cowley, T. Hsu, *Micron Microscopia Acta*, **1987,** *18*, 179.
[33] F. Tsai, J. M. Cowley, *Ultramicroscopy* **1992,** *45*, 43.
[34] T. Hsu, J. M. Cowley, *Ultramicroscopy* **1983,** *11*, 239.
[35] N. Osakabe, Y. Tanishiro, K. Yagi, G. Honjo, *Surf. Sci.* **1981,** *109*, 353.
[36] N. Osakabe, Y. Tanishiro, K. Yagi, G. Honjo, *Surf. Sci.* **1980,** *97*, 393.
[37] J. M. Cowley in *Surface and Interface Characterization by Electron Optical Methods* (Eds.: A. Howie, U. Valdre), Plenum Press, New York and London **1988,** p. 127.
[38] N. Yao, Z. L. Wang, J. M. Cowley, *Surf. Sci.* **1989,** *208*, 533.
[39] T. Hsu, J. M. Cowley in *The Structure of Surfaces* (Eds.: M. A. van Hove, S. Y. Tong), Springer-Verlag, Berlin **1984,** p. 55.
[40] Y. Tanishiro, K. Takayanagi, K. Yagi, *J. Microsc.* **1986,** *142*, 211.
[41] G. Lehmpfuhl, Y. Uchida, *Surf. Sci.* **1990,** *235*, 295.
[42] J. Liu, L. Wang, J. M. Cowley, *Surf. Sci.* **1992,** *268*, L293.

[43] H. A. Bethe, *Ann. Phys.* **1928,** *87*, 55.
[44] R. Colella, *Acta Crystallogr.* **1972,** *A28*, 11.
[45] P. A. Maksym, J. L. Beeby, *Surf. Sci.* **1984,** *140*, 77.
[46] A. Ichimiya, *Jpn. J. Appl. Phys.* **1983,** *22*, 176.
[47] S. Y. Tong, T. C. Zhao, H. C. Poon in *Reflection High Energy Electron Diffraction and Reflection Electron Imaging of Surfaces* (Eds.: P. K. Larsen, P. J. Dobson), Plenum Press, New York and London **1988,** p. 63.
[48] J. M. McCoy, P. A. Maksym, *Surf. Sci.* **1994,** *310*, 217.
[49] P. Goodman, A. F. Moodie, *Acta Crystallogr.* **1974,** *A30*, 280.
[50] J. M. Cowley, P. M. Warburton in *The Structure and Chemistry of Solid Surfaces* (Ed.: G. A. Somorjai), J. Wiley and Sons, New York **1967,** p. 6-1.
[51] L.-M. Peng, J. M. Cowley, *Acta Crystallogr.* **1986,** *A42*, 552.
[52] Y. Ma, L. D. Marks, *Microsc. Res. Tech.* **1992,** *20*, 371.
[53] Z. L. Wang, P. Lu, *Ultramicroscopy* **1988,** *26*, 217.
[54] K. Yagi, Y. Akira, I. Homma, *Microsc. Res. Tech.* **1992,** *20*, 333.
[55] P. A. Crozier, M. Gajdardziska-Josifovska, J. M. Cowley, *Microsc. Res. Tech.* **1992,** *20*, 426.
[56] Z. L. Wang, J. Bentley, *Ultramicroscopy* **1993,** *51*, 64.
[57] F. Tsai, J. M. Cowley, *Ferroelectrics* **1993,** *140*, 203.
[58] N. Osakabe, N. Endo, J. Matsuda, A. Fukuhara, A. Tonomura, *Phys. Rev. Letts.* **1989,** *62*, 2969.
[59] H. Banzhof, K. H. Herrmann, H. Lichte, *Microsc. Res. Tech.* **1992,** *20*, 450.
[60] K. Yagi, *J. Appl. Crystallogr.* **1987,** *20*, 147.
[61] N. Yao, J. M. Cowley, *Ultramicroscopy* **1989,** *31*, 149.
[62] L. Wang, J. Liu, J. M. Cowley, *Surf. Sci.* **1994,** *302*, 141.
[63] N. Yao, J. M. Cowley, *Ultramicroscopy* **1990,** *33*, 237.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

1.3 Electron Energy-Loss Spectroscopy Imaging

1.3.1 Introduction

Electron energy-loss spectroscopy (EELS) measures the spectral distribution of energy transferred from a monochromatic incident electron beam into a specimen. Two main types of geometry are concerned: low energy primary beams reflected by solid surfaces, and high energy primary beams transmitted through thin foils or at glancing incidence along surfaces. In the first case, using monochromators, the investigated excitation spectrum covers the millielectronvolt to electronvolt range (which is equivalent to the infra red domain of the electromagnetic spectrum) and is largely a vibrational spectroscopy. In the second case, corresponding to the electron microscopy situation, the inelastic scattering events involved encompass a whole range of excitations from the electronvolt to the kiloelectronvolt range, which is equivalent to the visible to soft X-ray domain. It is, therefore, a spectroscopy of electron states, related to the excitation of the valence and conduction electrons in the low energy-loss domain, i.e. from 1 to 50 eV, and of the atomic-core electrons, in the high energy-loss domain from about 50 eV up to one- or several-thousand of electronvolts.

When performed in the context of an electron microscope column, the EELS technique obviously introduces a new dimension to electron microscopy. As a consequence of the elemental specifity of the signal delivered, EELS is now well established as an essential component of analytical microscopy, extending the accessible spatial resolution to the subnanometer level. Furthermore, it is largely accepted that the access to the information stemming from inelastic processes, in parallel to that contained in the elastic ones, constitutes an essential step forward in promoting electron imaging and diffraction as real quantitative tools. This chapter describes the present state of instrumentation and methodology accessible in the field of EELS imaging and emphasizes the great diversity of potential fields of application through a selection of recent investigations. For a more complete knowledge of the impact of EELS as a spectroscopy of electron states (compared to photon beam techniques), the reader is advised to consult other reviews [1–5].

1.3.2 Instrumentation and Methods

1.3.2.1 An Historical Survey

One can trace back the origin of the use of electron energy losses as a micro-analytical technique to the early work of Hillier and Baker [6], five decades ago. However, it only became used practically in the 1960's, as a consequence of the progress achieved in realizing and coupling well-adapted analyzers and filters to an electron microscope column. Two major approaches (scanning or fixed beam) have been explored, leading to energy-analyzing and energy-selecting microscopies.

In the energy-analyzing mode, a spectrometer transforms a point on the object into an EELS spectrum. From the prototypes designed by Wittry [7] to be introduced at the bottom of a conventional transmission electron microscope (CTEM) column, and by Crewe et al. [8] as an essential part of their first scanning transmission electron microscope (STEM) microscope, it has evolved into the present Gatan commercially available attachment realized by Krivanek and Swann [9]. It is basically made of a magnetic sector with tilted and curved entrance and exit surfaces of the pole pieces in order to deliver a second-order aberration-corrected image of the spectrometer object function. This combination of a STEM column with such an EELS spectrometer is particularly well suited to recording EELS spectra from local areas. When governed by the intensity distribution within the probe issued from a field emission source, the accessible spatial resolution can be as small as a fraction of a nanometer. Furthermore, the upgrading of the initial serial detection system (which comprised a slit and a scintillator-photomultiplier device) by a parallel array of typically 1000 diodes which measure in parallel all the spectral channels, has been responsible for a major breakthrough in terms of detection efficiency [10]. Such a combination has demonstrated a routinely accessible energy resolution of about 0.5 eV on core edges and identification capabilities approaching the single atom level, when coupled to a FEG STEM [11, 12]. Figure 1 shows the basic configuration and major components

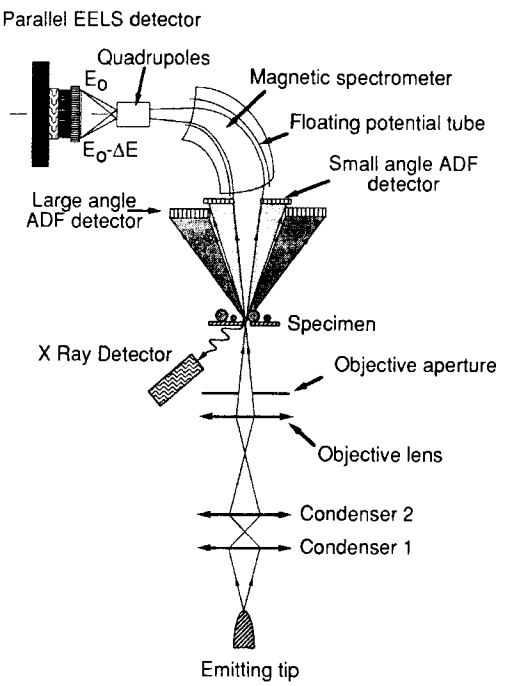


Figure 1. Typical example of the STEM + PEELS configuration for point analysis. It offers the simultaneous capability of visualizing the object topography through the annular detectors (small-angle for diffraction contrast, large-angle for Z-contrast) and of analyzing with the magnetic spectrometer the energy-loss distribution of the forwardly transmitted beam. (ADF = annular dark field.).

of the dedicated STEM equipped with an EELS spectrometer and a parallel detector array, in operation in Orsay. Although the performance of the diode array has not yet been optimized, it constitutes the best approach presently available for the acquisition of single EELS spectra originating from well-defined areas on the specimen, and a variety of processing techniques have been developed to extract the useful information from these spectra [13, 14]. An alternative solution to the magnetic sector is the Wien filter. It has been implemented on a dedicated VG STEM by Batson [15], and this unique system on an upgraded microscope provides the best presently attainable performance in terms of both energy resolution (of the order of 0.15 eV) and spatial resolution (down to 0.2 nm). However, it is not suitable for routine use and has not been duplicated by any other group or manufacturer.

In the energy-selecting mode, an imaging stage is added to the spectrometer, transforming the spectrum behind the energy-selecting slit into an image corresponding to a specific energy loss. This approach has been pioneered by Castaing and Henry [16], who produced the first high-quality energy-filtered images using a magnetic prism–electrostatic mirror device incorporated in the imaging part of a microscope column. Obviously, their realization has offered quite new possibilities to conventional transmission electron microscopy (TEM), such as a type of 'colour' microscopy, as different images corresponding to different types of inelastic interaction can then be obtained from a selected area. The first microanalytical applications followed rather quickly: El Hili [17] used the change of plasmon energy in different metals and alloys and Colliex and Jouffrey [18] showed energy-filtered images with electrons corresponding to a specific core loss.

In spite of these promising results, the first commercial attempt at producing an energy-filtering microscope using the Castaing–Henry design, by the French company Sopelem, has not been a clear success. It took nearly 20 years before Zeiss introduced to the market its 902 microscope, using the same concept, then its fully magnetic substitute, known as the 'omega filter', in the 912 machine. As a matter of fact this second version also had its origin in Castaing's group [19], before being developed by Zanchi et al. [20] and optimized by Rose and coworkers [21, 22].

However, these filters are only available on dedicated instruments and cannot be retrofitted to existing microscopes. An alternative possibility, first explored by Shuman and Somlyo [23], consists of an imaging spectrometer where a simple magnetic sector acting as an analyzer is followed by an electron optics transfer system which displays the energy filtered image on a two-dimensional (2D) recording medium. The very sophisticated system designed, produced and tested by Krivanek and coworkers [24, 25], known as the Gatan imaging filter (GIF), can be adapted to any TEM column (even on megavolt instruments) and provides in a quite user-friendly configuration energy-filtered images with atomic resolution. Figure 2 compares the two approaches to the fixed-beam energy filtering presently commercially available, the Gatan 1-6-6 filter and the Zeiss Ω filter.

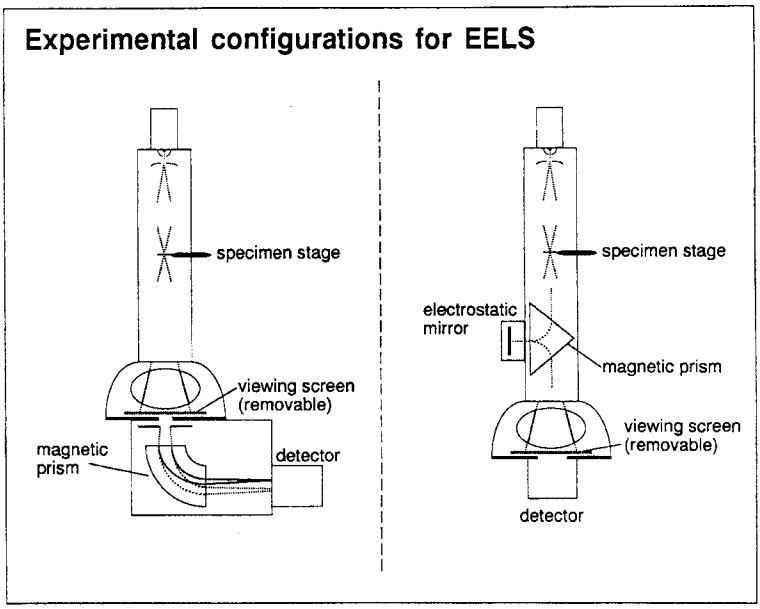


Figure 2. Typical examples of the incorporation of filtering devices in a TEM column for energy selected imaging (ESI). A change in the excitation of the first intermediate lens replaces the diffraction pattern in the entrance object plane of either filter by an image of the specimen, and replaces the final energy-filtered image on the detector by an energy-filtered diffraction pattern.

1.3.2.2 A New Dimension in EELS Imaging: From Point Analysis and Energy-Filtered Images to Spectrum Imaging

The introduction of the energy-loss parameter (ΔE) adds a new dimension of information to the current 2D spatial information provided for a given pixel position (x, y) by any type of scattering responsible for the contrast observed in a conventional electron micrograph. One can then measure the intensity $I(x, y, \Delta E)$ within an elementary volume defined within a three-dimensional (3D) space with two axes related to the position and one to the spectral information. For any of these elementary volumes, a given dose of electrons $(J \times \tau)$ incident on the entrance surface of the specimen, is required in order to produce a measurable signal with a given signal-to-noise ratio, which depends practically on the physical parameters (thickness and composition) of the specimen and of the investigated spectral channel.

Basically, a spectrometer in a STEM mode with parallel acquisition provides a full spectrum $I(\Delta E)$ covering n energy-loss channels for each probe position, i.e. for each pixel defined by its coordinates (x, y) on the specimen within the time interval Δt_1 set by the dose requirements (see Fig. 3a). On the other hand an energy filtering microscope (EFTEM) provides a complete 2D image made up of $N \times N$ pixels, using only the electrons contained within an energy band defined by the selection slit (see Fig. 3b). Similar criteria of signal-to-dose ratio apply for the evaluation of the image recording time Δt_2. To obtain the same intensity $I(x, y, \Delta E)$ on a given area element within a given energy-loss channel, it implies, as stated above, that $J_1 \times \Delta t_1 = J_2 \times \Delta t_2$, all

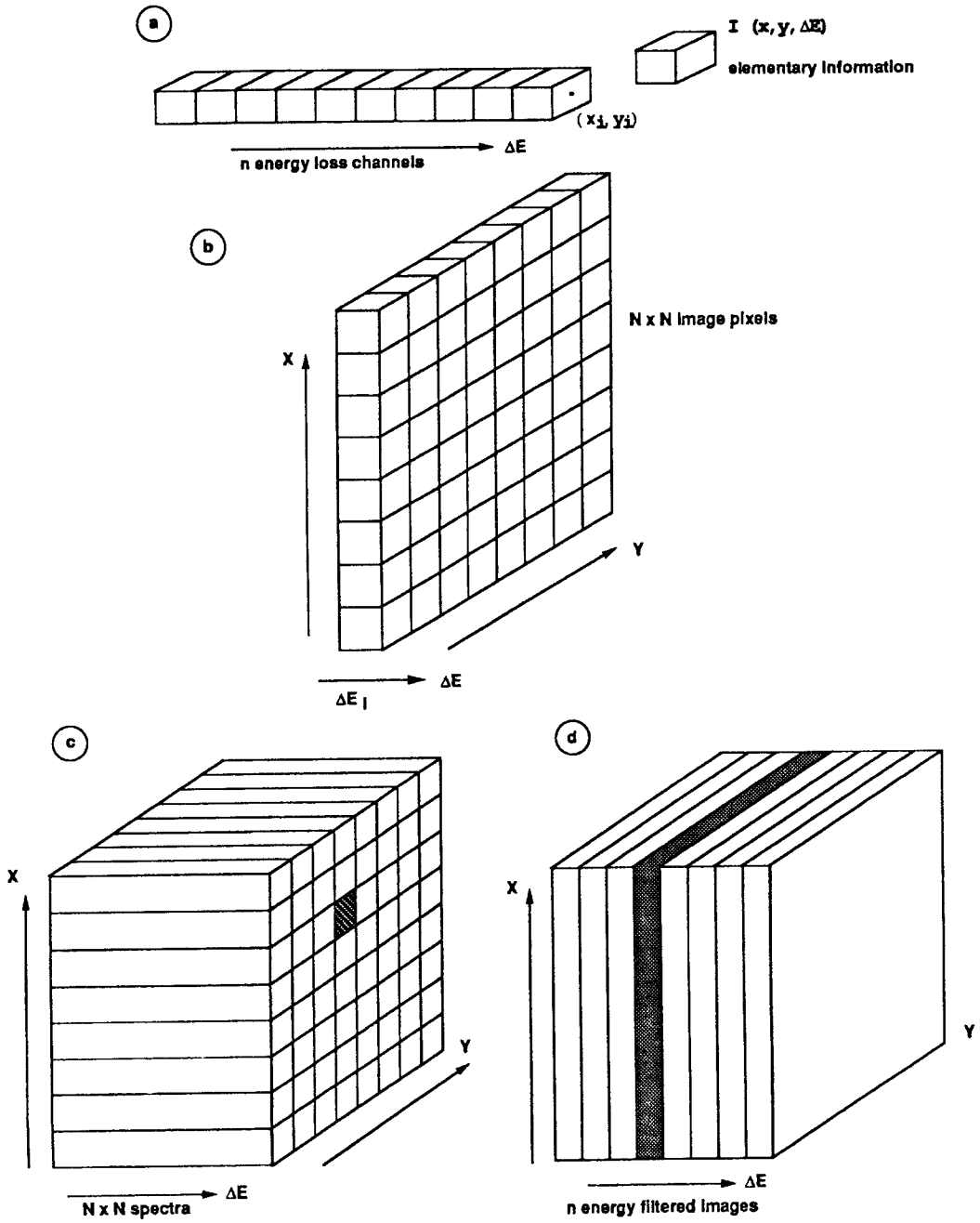


Figure 3. Definition of: (a) the parallel acquired EELS spectrum in a STEM configuration; (b) the energy filtered image in an EFTEM configuration; the image-spectrum acquired either as a 2D collection of parallel EELS spectra (c) or as a sequence of energy filtered images (d).

other parameters governing the measured signal being equal. The ratio between the recording times Δt_1 and Δt_2 depends on the ratio of the primary flux of electrons available to the specimen. Practically, this is about 10^2–10^3 times higher in a nanoprobe delivered from a field emission source in a STEM than in a submicroprobe delivered from a LaB6 filament in an EFTEM, and $\Delta t_1/\Delta t_2 = a$ is of the order of 10^{-2}–10^{-3}.

A few years ago, Jeanguillaume and Colliex [26] introduced the concept of the spectrum-image in EELS digital acquisition and processing. It was originally stimulated by the access to parallel EELS recording devices in the STEM and by the availability of computers with rapidly increasing capacities for data storage and handling. As a matter of fact, this concept can be applied to any case where spectra can be acquired as a function of a probe location, such as energy-dispersive X-ray (EDX) or Auger analysis under a primary electron beam, inelastic tunnelling spectroscopy under the scanning tip, and secondary ion mass spectrometry (SIMS) by sputtering under an ion probe. A spectrum image is a 3D ensemble of numbers $I(x, y, \Delta E)$: the first two axes correspond to the (x, y) position on the specimen, similarly to any image and the third axis is associated with the energy-loss dimension (ΔE). As shown in Figs. 3c and d, the spectrum image can be built with the basic entities provided by the STEM + PEELS analyzing microscope (Fig. 3a) and by the EFTEM filtering microscope (Fig. 3b). It can be described either as a collection of $N \times N$ spectra acquired while scanning the STEM probe over the specimen surface, or as a stack of n filtered images recorded successively while ramping the energy-loss value of the electrons transmitted through the selection slit. The total time T needed for recording this amount of $N \times N \times n$ values is, in the first case, $N \times N \times \Delta t_1$ and, in the second case $n \times \Delta t_2$, but the total dose D received by the elementary area $\Delta x \, \Delta y$ is proportional to $J_1 \times \Delta t_1$ for the STEM and to $J_2 \times n \times \Delta t_2$ for the EFTEM. This leads to:

$$T_1/T_2 = N \times N \times a/n \quad D_1/D_2 = 1/n \quad (1)$$

Consequently, one can summarize this comparison between the two approaches, as follows:

(i) The parallel EELS + STEM mode is always better in terms of the dose required, because of the intrinsic superiority of the technique in recording all energy-loss channels in parallel, while the EFTEM technique requires as many images (and irradiations) as the number of energy-loss channels required.

(ii) The comparison is not as obvious when considering the total time required for the acquisition. For small numbers of pixels and high numbers of loss channels, the PEELS + STEM is superior, while for large N and small n values the EFTEM provides better results. Of course, for a single filtered image, such as the selection of the elastic image, filtering devices have to be recommended.

In any case, however it has been acquired, the great advantage of the spectrum-image is that it can be processed a posteriori, offering access to the exact relationship between the spatial coordinate and the spectral feature. For instance, elements in unexpected locations may be found without any a priori operator

decision regarding where to locate the beam for data collection. Furthermore, it allows summation of spectra within segmented areas, thus improving the statistical significance of the data. On the other hand, the major limitation is the large data capacity required (32 Mbyte for a $128 \times 128 \times 1024 \times 16$ bit spectrum-image) and the time needed for access and manipulation of data in any type of off-line processing. But this difficulty will surely gradually vanish as faster and cheaper computers become available.

The hardware and software required for implementing spectrum-imaging techniques have been used over the last few years by a number of groups. In particular, Hunt and Williams [27] have published the first complete description of a system installed on a VG HB501 at NIH, Bethesda, MD, and on a Philips 430 TEM/STEM at Lehigh University, Pennsylvania, USA, illustrated with results obtained in both life and materials sciences [28, 29]. On the other hand, the access to image-spectra using a fixed-beam filtering microscope has been demonstrated by Lavergne et al. [30]. The number of facilities that have developed their own equipment has increased during the past year, and the examples given in the following sections of this Handbook have been acquired and processed using the system described previously [31]. The system consists of a combination of a FEG VG STEM with a Gatan 666 PEELS in which the array of photodiodes has been modified to reduce the read-out noise. It makes profiling and mapping of EELS spectra achievable with (sub)nanometer spatial resolution. The experimental parameters routinely employed are summarized in Table 1. For reasons already discussed (limited number of pixels), the present installation is particularly well suited to the investigation of line spectra, that is, to one-dimensional (1D) spatial data acquired when scanning the probe along a given line on the specimen.

Table 1. Parameters currently used for the acquisition of spectrum lines and spectrum images with a PEELS + STEM configuration

	Probe size (nm)	Typical extent of the used signal (nm)	No. of pixels involved	Pixel step on the specimen (nm)	Average recording time per spectrum
Spectrum-line					
Low losses (plasmons, dielectric coefficients)	1–2	>1–2	64–256	0.5–1	25–250 ms
Elemental mapping from core-losses	0.5	<0.5	32–128	0.3–0.5	1 s
Fine structures on core-losses	0.5	<0.5	32–128	0.3–0.5	1–5 s
Spectrum image					
Elemental mapping from core-losses	1–2	<0.5	32×32 or 64×64	0.5–1	0.1–1 s

Another interesting possibility is the access to time-resolved spectroscopy ('chronospectroscopy' [32]), in which case the incident beam is maintained fixed on the specimen and sequences of spectra are acquired at fixed time intervals and reflect in particular the change of the specimen as a function of the dose. This possibility has proved to be of great use when investigating beam-sensitive specimens, in which case it is possible to estimate the zero-dose spectrum by techniques of back-extrapolation.

1.3.3 Understanding and Extracting the Information Contained in an EELS Spectrum

1.3.3.1 Anatomy of an EELS Spectrum

A typical EELS spectrum [1–5] exhibits the characteristic features shown in Fig. 4: a zero-loss peak; a major contribution (the low-loss region) extending from

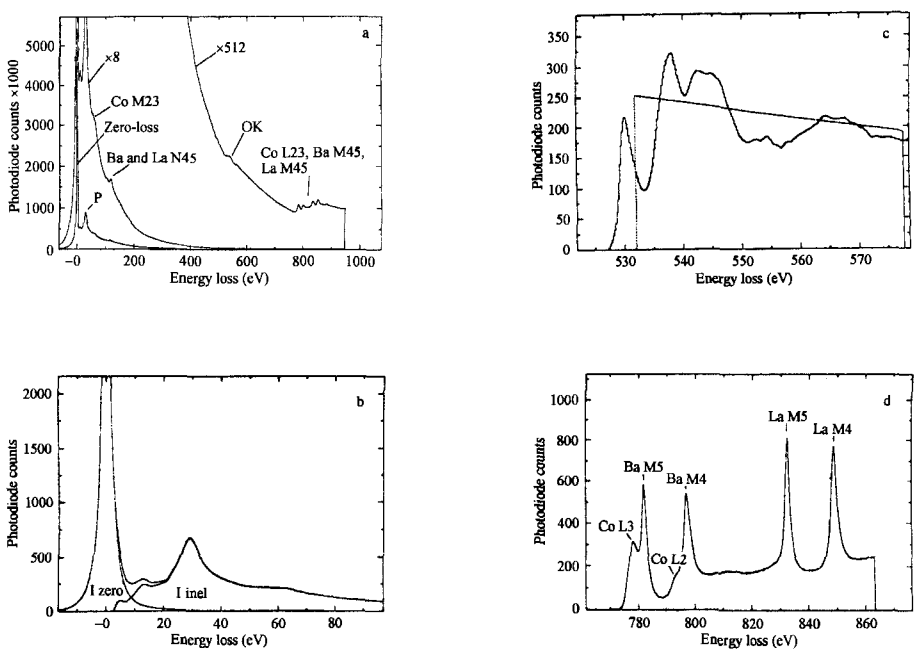


Figure 4. An EELS spectrum acquired on a thin foil of nominal composition $LaBaCo_2O_{5+\delta}$ (specimen courtesy of F. Studer, Caen). (a) The whole spectrum extending over 1000 eV covers four orders of magnitude in intensity for comparison with the non-saturated zero-loss peak and the details of the overlapping Co, Ba and La edges around 800 eV loss. (b) The tail of the zero-loss peak and the onset of inelastic scattering, which can be used to evaluate the local thickness in terms of the inelastic mean free path (t/λ_{in} is of the order of 0.63 in this present spectrum). (c) The oxygen K edge after background subtraction and multiple loss deconvolution, compared with an atomic calculation of the relevant cross-section (using the Sigmak program of Egerton). Note how the fine structures modulate the calculated cross-section over the ELNES range. (d) The overlapping white lines and edges for CoL_{23}, BaM_{45} and LaM_{45}.

about 5 to 50 eV and corresponding to the excitation of electrons in the valence band and low-lying levels; and, superimposed on a continuously decreasing background, a succession of element-specific edges at increasing energy losses. All these features contain useful information, which in certain cases can only be reached after some lengthy procedure. Briefly, the low-loss structures reflect a mixture of collective ('plasmon') and interband processes, i.e. some kind of average electron properties. For example in a material exhibiting quasi-free-electron gas behavior, the energy of the plasmon peak is proportional to the mean density of electrons.

On the other hand, the core-loss features that appear between 100 and 1000 eV are chemically representative and constitute the major route to nano- and micro-analytical applications. However, it is obvious that all visible edges do not display similar shapes, their general behavior being imposed by atomic considerations. If the

Table 2. Summary of the different types of information accessible through a detailed analysis of an EELS spectrum

Spectrum domain	Information accessible	Required processing technique	Field of application
Whole spectrum and zero loss in particular	Thickness, total inelastic scattering	Measurement of unsaturated zero-loss peak	Very general
Low-loss region	Average electron density	Measurement of plasmon line properties	Microanalysis in metallic alloys
Low-loss region	Interband transitions/joint density of states	Kramers–Kronig transformation; calculation of dielectric constants (critical-point modeling)	Optical and transport properties; comparison with VUV spectra; intergranular van der Waals forces
Low-loss region	Interface/surface properties	Study of the interface plasmon modes	Interface and boundary structure and chemistry
Core-loss region	Qualitative and quantitative elemental analysis	Measurement of core-edge weight	Nanoanalysis of any type of material
Core-loss region	Site symmetry; bonding type; bond lengths	Analysis of core-ELNES; comparison with fingerprints; molecular orbital, multiple scattering or band structure calculations	Site-selected valence state; charge transfer; bonding and structural environment
Core-loss region	Radial distribution function	Analysis of core-EXELFS	Site-selected crystal coordination
Core-loss region	Density of holes on local states	Measurement of white-line intensities	Electron configuration in intermetallics, insulators, and superconductors; magnetic properties

ELNES, energy loss near edge structure; EXELFS, extended electron energy loss fine structure; VUV, vacuum ultraviolet.

oxygen K edge exhibits a rather typical sawtooth profile modulated by fine structures, the cobalt L_{23} edge and the Ba and La M_{45} edges appear as sharp lines ('white lines') followed by a step function. These are due to strong transitions from the initial levels with p- and d-type symmetry to unoccupied bound states of 3d and 4f symmetry, respectively, reflecting (at least indirectly) the population of d or f holes on the excited atom. Table 2 summarizes the different types of information accessible through an analysis of the different spectral regions.

1.3.3.2 Methods Available for Data Processing and Specific Information Extraction

In parallel with the continuous improvement in understanding the physical content of EELS spectra, methods have been elaborated for processing spectra in order to extract from them the different types of specific information listed in Table 2 [33–35]. The most important methods are those dealing with quantitative elemental analysis. The first step consists in evaluating the local thickness and the importance of multiple scattering events. Assuming a Poisson distribution for the occurrence of independent scattering events

$$P_n(t) = \frac{1}{n!}\left(\frac{t}{\lambda_i}\right)^n \exp\left(-\frac{t}{\lambda_i}\right) \qquad (2)$$

one calculates

$$t/\lambda_i = \log(I/I_0) \qquad (3)$$

where the thickness is scaled to the inelastic mean free path (λ_i) corresponding to all possibilities of an inelastic event, whatever may be the scattering angle β or the energy loss ΔE. The information is then contained in the ratio between the intensity of the zero-loss peak (I_0) compared to that of the total energy loss spectrum with the zero-loss peak included (I). As soon as the ratio t/λ_i is higher than about 0.5, deconvolution techniques must be employed in order to recover the single scattering signals.

The basic formula for quantitative elemental analysis with a core-loss relates the measured signal (S) to the number of atoms generating it:

$$S = I_{\text{inc}} N \sigma \tau \qquad (4)$$

where I_{inc} is the incident beam current, N is the number of atoms per unit specimen area giving rise to the measured signal, σ is the relevant cross-section, and τ is the counting time.

Egerton [34] has shown that a good approximation, when measuring the signal for a given angle of collection (β) and integrated over an energy window (Δ) over threshold, is to introduce in the above formula the low-loss spectrum $[I_0(\beta, \Delta)]$ and the cross-section $[\sigma(\beta, \Delta)]$ measured under the same conditions as the signal $[S(\beta, \Delta)]$.

There are two types of problem associated with achieving a good quantitative analysis: the measurement of the characteristic signal S as discriminated from the background, and accessing reliable cross-section values. The extraction of the characteristic edge implies background modeling over a fitting window below threshold and extrapolation of this model curve under the signal before subtraction (Fig. 5a). The most generally used model is of the $A \cdot \Delta E^{-R}$ type. It is a simple method of broad applicability and requires no

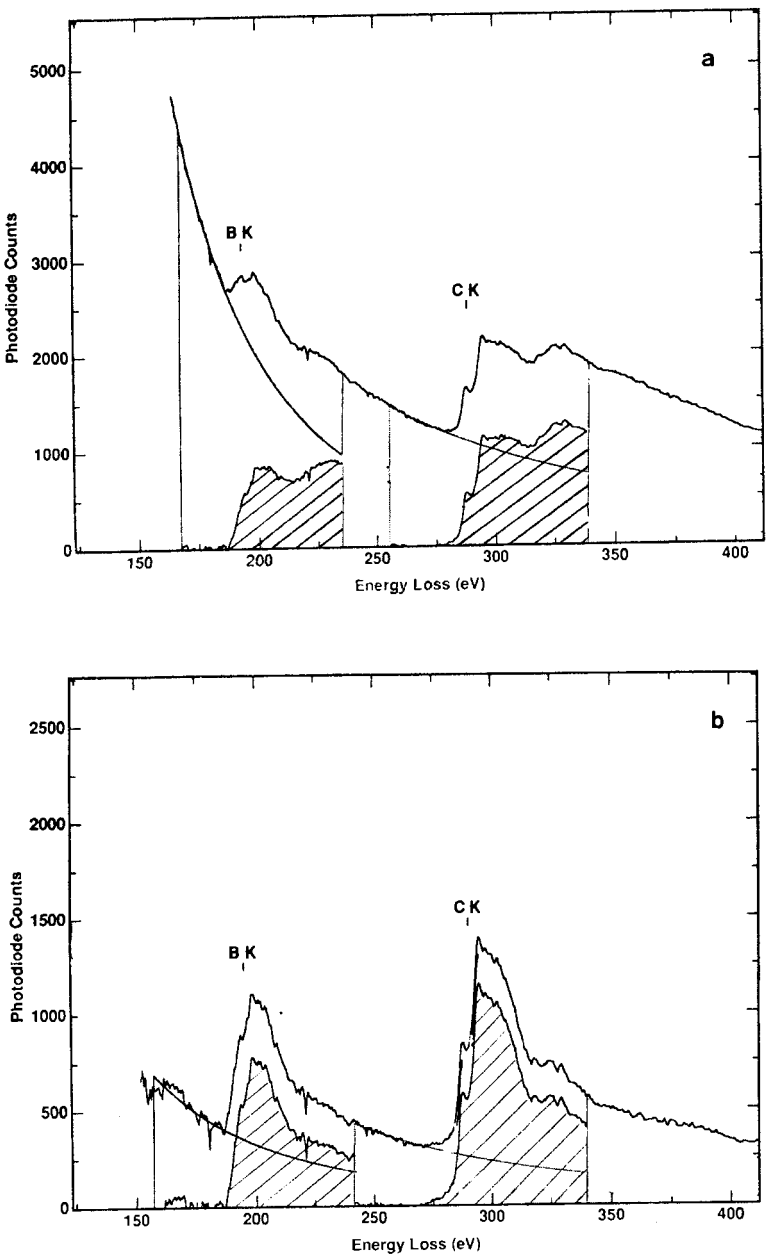


Figure 5. Quantitative analysis of a BC_3 thin foil with $t/\lambda_i = 0.87$, using the background extrapolation technique under the BK and CK edges: (a) original spectra, (b) deconvoluted spectra. Using the Sigmak cross-sections, one finds in the first case B/C = 0.32 and in the second case B/C = 0.34.

other information. However, it may fail because the power-law model is not well adapted in the low-loss regime (i.e. up to 100–150 eV). Furthermore, this method cannot be used where there are overlapping edges, and the accuracy of the measurement of the signal depends on the quality of the fit [estimated by an h factor (following Egerton) or by the reduced χ^2 term].

The cross-sections are satisfactorily estimated using a hydrogenic model for the K and L shells, and the relevant values of $\sigma(\beta, \Delta, E_0)$ calculated by Egerton [36] can be found in the routinely available quantification procedures such as the ELP software available from Gatan. For the outer shells of M, N or O type, the calculations are much more complex and require an Hartree–Slater description [37]. The values calculated by Rez are now available in the latest version of the ELP software. The reliability of these calculated values has been tested on reference specimens by many authors, in particular by Hofer [38]. Hofer has shown that for the K and L shells the accuracy of the calculations is of the order of ±5% for most light-element edges up to the transition metal series. However, for heavier elements the edges involved require the use of the more complex calculations and discrepancies up to 40% have been found when comparing calculated values with measured ones.

When one is only interested in measuring relative concentrations of different elements (A and B), the above formula can be simplified as:

$$\frac{N_A}{N_B} = \left[\frac{S_A(\beta, \Delta)}{S_B(\beta', \Delta')}\right]\left[\frac{\sigma_B(\beta', \Delta')}{\sigma_A(\beta, \Delta)}\right] \qquad (5)$$

where the ratio of cross-sections corresponds to a scaling k factor (following the terminology used in EDX analysis of thin foils), reduced to the ratio of the probability of occurrence of the edges involved.

For specimen thicknesses ranging from about $0.5\lambda_i$ to $1.5\lambda_i$, the same formula can still be applied, by replacing the experimental signals with the deconvoluted ones. Two alternatives can be used practically: (i) the Fourier-log method and (ii) the Fourier-ratio method. Both methods require a knowledge of an unsaturated low-loss spectrum from the same area. The first technique applies to the whole spectrum from zero loss to core losses, background included. The second method only requires independently recorded core-loss and low-loss spectra. Although it may introduce more noise, the first technique is generally recommended. Figure 5b shows the quantitative analysis applied to the deconvoluted spectra obtained from the original shown in Fig. 5a.

1.3.3.3 Alternative Approaches using Multiple Least-Squares Techniques

In many situations, the above standard technique consisting of background modeling, extrapolation, and subtraction is not adapted. One can quote the case of overlapping edges, the need for evaluating the weight of different components within an edge, and, in general, the detection of very small concentrations. In the latter case, the signal appears only as a weak modulation over an intense background and its measurement can be made more difficult when using a parallel recording device composed of a collection of individual detection units due to the inhomogeneities between the detection properties of the different detectors. For these reasons, new tools for recording and processing EELS data have been introduced. To reduce the influence of the variations of read-out and dark-current noises in the different channels of

a diode array used for parallel recording, the use of difference spectra (first or second difference) is a simple approach which requires a double or triple exposure of the same spectrum shifted by a given energy increment (δ) and covering different detector distributions. This method has proved to be quite efficient [13, 14].

In order to process all these different types of EELS spectra, multiple least-squares (MLS) fitting techniques are being rapidly developed [39–41]. These methods search for the combination of reference signals which best reproduce an experimental curve, encompassing both the background and the characteristic edges. The procedure is quite versatile and can be used for different purposes. It implies the use of reference spectra, either calculated or previously obtained from standards. The thickness effect and its associated multiple scattering events can be accounted for by convolution techniques. The major limitation in using MLS fitting techniques for EELS spectra, as compared to the simpler case of processing EDX data where all characteristic lines can be fitted with similar profiles, is due to the fact that the detailed shape of the edge exhibits variable fine structure, depending on the local environment of the atom. For example, one cannot use cation edges recorded from metals when the cation is inserted in an ionic compound.

These limitations being kept in mind, its application for quantitative analysis technique has been demonstrated successfully by Leapman [13] on different examples. In particular, the detection limits have been improved considerably when the magnitude of the derivatives of the cross-section is high (in particular when the edge exhibits intense white lines at threshold). Figure 6 shows an example from Leapman and Newbury [42], extracted from a test experiment performed on a glass specimen containing 66 elements, and in particular transition and rare earth elements at trace concentrations. When considering the quality of the fit for the La M_5–M_4 lines, one can estimate the detection limit for such an element under the experimental conditions used to be of the order of 10 ppm. One further parameter controlling the detectability limit is then the edge shape:

(i) 3d transition elements and preceding alkaline earth L_{23} edges (from Ca to Ni), as well as 4f lanthanides and preceding alkaline earth M_{45} edges (from Ba to Yb) can be detected in the 10–100 ppm atomic concentration range;

(ii) light elements identifiable through their K edge (Li to P) or their L_{23} edge (Mg to Cl) and 4d transition elements identifiable through their M_{45} edge (Sr to Rh) can usually be detected in the 100–1000 ppm range; and

(iii) for all elements with no clear discontinuity at threshold (Ga, Ge, As,..., Sb, Te, W,...) it seems reasonable to estimate a normal detection limit around an atomic concentration of 1%.

Another important field of application for these fitting techniques is to obtain access to quantitative measurements of the different contributions within the fine structures of a given edge, i.e. the relative strength of a pre-peak on an O K edge such as the one shown in Fig. 4c, or the weight of the unoccupied d holes as compared to that of transitions to continuum states $[(I_{L3} + I_{L2})/I_{cont}]$ on an L_{23}

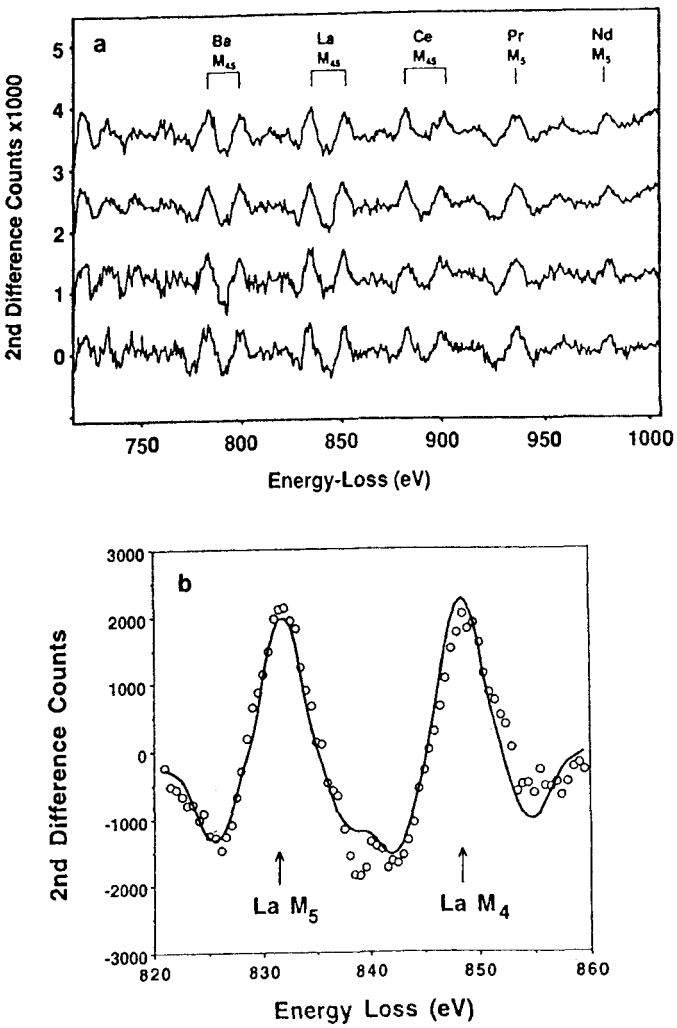


Figure 6. Detection of elements in the concentration range 10–100 ppm in a reference glass, using the MLS fitting technique on a second difference spectrum. (a) Four different spectra exhibiting the M_{45} white lines for the light rare-earth elements from 750 to 1000 eV. (b) Best fit for the detection of the La signal corresponding to a nominal concentration of 76 ppm. (Courtesy of Leapman and Newbury [42].)

edge of a transition metal. Figure 7 demonstrates how the MLS fitting technique described by Manoubi et al. [41] can be used to evaluate separately the weight of the two white lines (I_{L3} and I_{L2}) and of the transitions to continuum states when trying to simulate an experimental Fe L_{23} edge as a sum of two Lorentzian curves for the white lines and of a Hartree–Slater cross-section in order to account for transitions to continuum states.

1.3.4 Applications of EELS Imaging

As described above, EELS spectroscopy and imaging techniques provide a great variety of approaches, adding a new dimension to TEM, that is, access to more refined quantitative or analytical microscopies (see, for example, Reimer et al. [43]). In this Section my intention is to illustrate some recent trends in the use

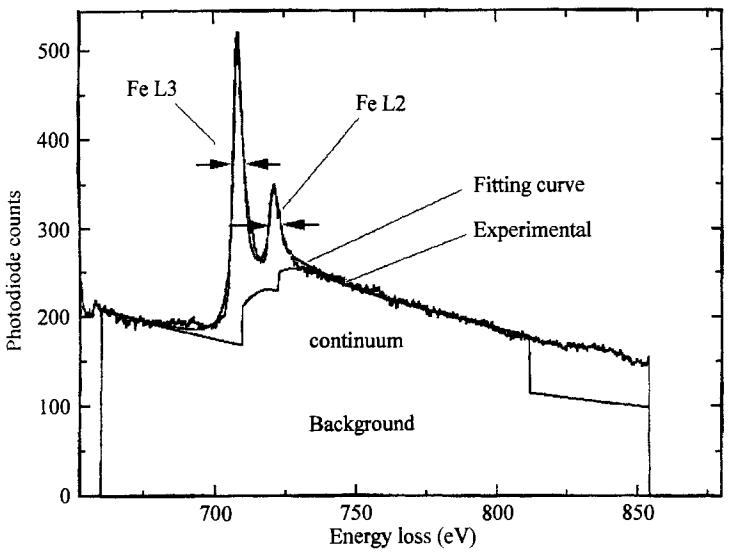


Figure 7. Use of a MLS fitting technique to measure the relative importance of different components in the fine structure of an FeL_{23} edge. The specimen is a transverse section of an Fe/Fe_xGe_{1-x} multilayer. An accurate measurement of the weight of the white lines and of their ratio is required for investigating the distribution of the electron states and the magnetic moment on the excited Fe atom. (Study performed in collaboration with K. Suenaga.)

of an imaging EELS spectrometer (either of the filter type or as a STEM + PEELS combination) in the field of materials science. Applications cover situations where the analysis is performed over many pixels, extending from a line profile of typically 100 points of analysis, to a full image comprising typically 1000 × 1000 pixels. As already explained, in the first case it is easy to record and process many spectral channels for each pixel, while in the latter case one can practically record only a limited number of images.

1.3.4.1 Advances in EELS Elemental Mapping

In essence, one makes a map of the spatial origin of chemically significant signals such as characteristic core edges. It is demonstrated here how the results are more quantitative when one has access to a greater number of energy loss channels for each pixel. The characteristic signal is always superimposed on a non-characteristic background. It is therefore obvious that with a single energy-loss image it is impossible to attribute a varying contrast to a change in chemistry. On the other hand, when processing a single EELS spectrum, it has been pointed out that the highest accuracy can only be obtained when the whole spectrum is known. In the case of moderately thick specimens, this includes the low-loss domain which is needed for making deconvolution corrections. In intermediate cases, the major problem is to discriminate changes in contrast due to compositional changes from those due to variations in thickness or orientation.

The first real progress in this direction was made when Jeanguillaume and co-workers [44, 45] proposed and realized the three-window technique in which two energy-filtered background images are acquired below the edge and one above the edge. An extrapolated background

image is then calculated, independently for each pixel, and subtracted from the image above the ionization threshold, thus providing a net image which can be considered as an elemental map. This technique avoids the risk of artifacts but is not strictly quantitative. In fact, the results should be scaled to the normal unfiltered bright field image corresponding to the I_0 term in Eq. (3). Nevertheless, the technique is a clear improvement on the two-window method in which a post-edge image was subtracted from a 'scaled' pre-edge image. However, another type of two-window method, known as the 'jump-ratio map', has been proposed recently [46]. The elemental map is simply made by dividing a post-edge by a pre-edge image. Hofer et al. [47] have demonstrated (Fig. 8), that this method produces maps with minimum added noise and which are only weakly affected by diffraction contrasts arising in crystalline materials. However, without denying these successes, I would not recommend systematic use of this technique because the interpretation of the measured contrast in terms of elemental composition is far from quantitative.

On the contrary, it is clear that any increase in the amount of information available and used leads to better quantification of the data. Crozier [48] has pointed out that for large-area mapping of specimens with complex morphology and varying thickness, inspection of the t/λ map derived from simultaneously recorded unfiltered and zero-loss images is a good basis on which to define which pixels in an image may be strongly affected by plural scattering. In an effort to assess the qualities of the different software available for elemental mapping from a

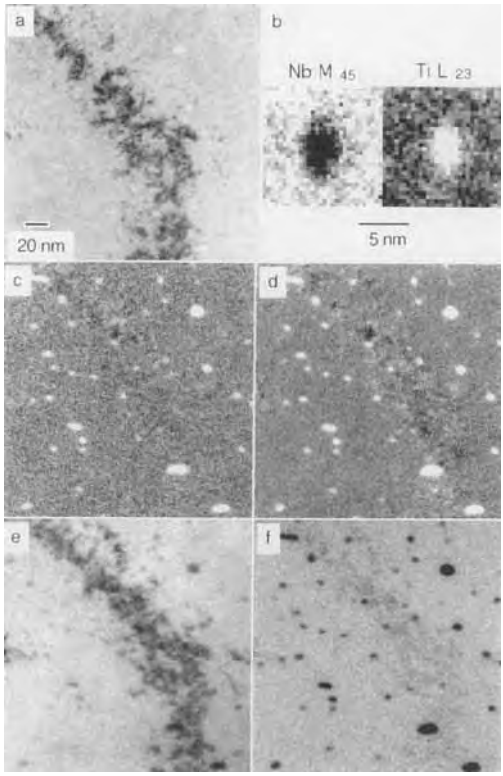


Figure 8. EELS mapping of nanometer-sized precipitates by processing several energy-filtered images acquired with an imaging filter attached on a 200-keV transmission electron microscope. The specimen was a thin foil of niobium containing TiO precipitates. The TiL_{23} edge was used for Ti mapping, and the niobium M_{45} edge was used for Nb mapping (b). The results obtained using the three-window technique (c, e), and the jump-ratio map (d, f). (a) Normal unfiltered TEM bright field micrograph showing the weak contrast of the precipitates close to a bend contour. (Data courtesy of F. Hofer et al.)

sequence of energy filtered images, Bonnet et al. [49] have pointed out the advantage of collecting an increased number of energy-filtered images both below and above the threshold in order to improve the quality of the fit and to enhance the signal per pixel. Furthermore, the signal-to-noise ratio (SNR) may benefit from local smoothing of the experimental data

before processing, as long as it does not destroy significantly the spatial resolution. A test specimen consisting of clusters comprising a small number of rare-earth or uranide atoms has been investigated in detail by Mory and Colliex [50] in order to estimate the limits of detection of the chemical mapping techniques. Introducing double exposure during the acquisition sequence (three pre-edge images and one post-edge image), these authors have demonstrated that the elemental identification of numbers of atoms below 10 can be performed with a good SNR. In a further step, it has been demonstrated that the implementation of linear optimal unbiased filters, such as those provided by geostatistical techniques (kriging procedure), could improve the SNR further, i.e. progressing toward single-atom identification [51].

The relationship between the different types of accessible performance (spatial resolution, minimum detectable number of atoms, and minimum detectable concentration) as a function of the physical constraints involved (SNR and cross-section for the used edge, and local specimen thickness) and of the experimental parameters used (primary dose and primary voltage) has been discussed by many authors, in particular by Colliex [52]. The examples cited above (concentrations of approximately 10^{-5} [42], and identification of a single atom [50, 51]) correspond to well-defined conditions in which the total current prevails in the first case and the local dose (need for a field emission source) in the second case. The intermediate case, where a filtering device on a TEM is used for the quantitative analysis of nanostructures, has been discussed theoretically and practically for typical examples (interfaces and nanotubes) by Berger et al. [53] and by Kurata et al. [54]. In both cases a subnanometer spatial resolution and a detection limit of a few atoms has been demonstrated, the signal-to-noise ratio being noticeably improved when summing the data over several equivalent pixels.

In all these cases, the edge is sufficiently clear so that the technique of extrapolating the background is sufficient to discriminate and measure the characteristic signal. However, as discussed above, in the case of single spectrum processing, there are many situations that deviate from this simple approach, in particular when the SNR is so low that the signal cannot be distinguished from fine structures on an edge at lower energy, or for very low concentration. It is then necessary to use some kind of MLS technique and the only possible means of making the calculations is to have access to a complete energy-loss spectrum for each pixel. The unique solution is the STEM + PEELS approach. Tencé et al. [55] have recently described some applications, using this type of routine for fitting the experimental spectrum acquired for each pixel with a linear combination of reference edges. When these references are recorded during the acquisition of other parts of the specimen, the technique is very powerful, as has been demonstrated by the analysis of an Au–Ni multilayer (Fig. 9). In this example, the situation is made very complex by the superposition of the edges involved (strong overlap between the Au O_{23} and N_{67} edges at 54 and 83 eV, respectively and the Ni M_{23} edge at 68 eV, requiring a MLS fit on the second-difference spectra) and by the spatial resolution required to identify a monolayer of Ni. Another highly

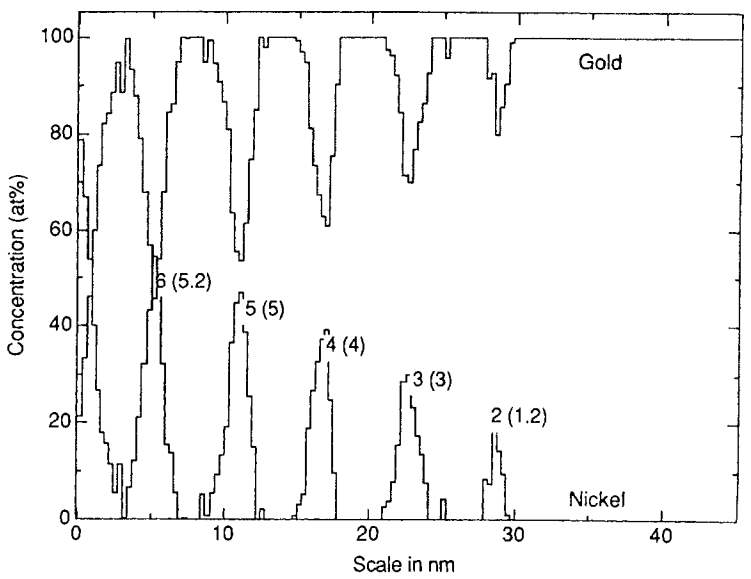


Figure 9. Chemical analysis of a transverse section of a Au–Ni multilayer obtained by processing a sequence of 64 second-difference spectra (energy-loss range 40–120 eV) acquired with step increments of 0.6 nm. It is important that the reference Au and Ni profiles are extracted from spectra recorded in the same sequence from areas containing only gold or nickel. Numbers in parentheses are proportional to the summed signal for each Ni layer; the other numbers correspond to the nominal values obtained during the molecular beam epitaxy deposition process. (Specimen courtesy of P. Bayle and J. Thibault.)

impressive study has been performed by Leapman et al. [56] concerning the mapping of Ca at very low concentration (a few 10^{-5} at. conc.) in a cryosection of cerebellar cortex. Here again MLS fitting techniques in the second difference mode were used on spectra summed within segmented areas defined on maps or images obtained at higher doses. These two types of result demonstrate clearly the trends to be followed for extracting the optimum information from EELS elemental analysis in situations at the limit.

1.3.4.2 Beyond Elemental Analysis: Mapping of EELS Fine Structures

Having recorded a series of spatially resolved EELS spectra, either in the spectrum-line or the spectrum-image mode, it is possible to extend the information extracted beyond simple elemental mapping or profiling by considering the complete spectral distribution over all relevant channels. Generally, the main interest is to identify which spectral feature is changing and at which probe position it is changing. Practically, the useful changes concern only a reduced number of energy-loss channels and a small number of pixels. Data-processing tools are being developed that will be able to answer these questions in an unbiased fashion.

The first solution is to calculate spatial-difference spectra, defined as the difference between spectra recorded at selected positions on the specimen; for example, on the matrix on each side of an interface and on the boundary itself [57]. This approach has been shown to be quite efficient in determining the bonding state and electronic structure of atoms at the interface itself, and thus opening up new avenues of investigating the atomic bonding

properties across interfaces [58, 59]. However, this approach must be performed manually, spectrum after spectrum, and is rather tedious and subjective.

As mentioned earlier, the present aim is to improve the quantitative aspects of profiling or mapping EELS fine structures, and consequently any associated information such as bonding type, valence state, d-state occupancy, and site symmetry. The means of achieving this involves the use of MLS fitting techniques, the preliminary step being to define or select templates of the fine structures which are to be searched along the line scan. This definition can be achieved a priori, relying on the knowledge of the specimen and of the problem (for instance

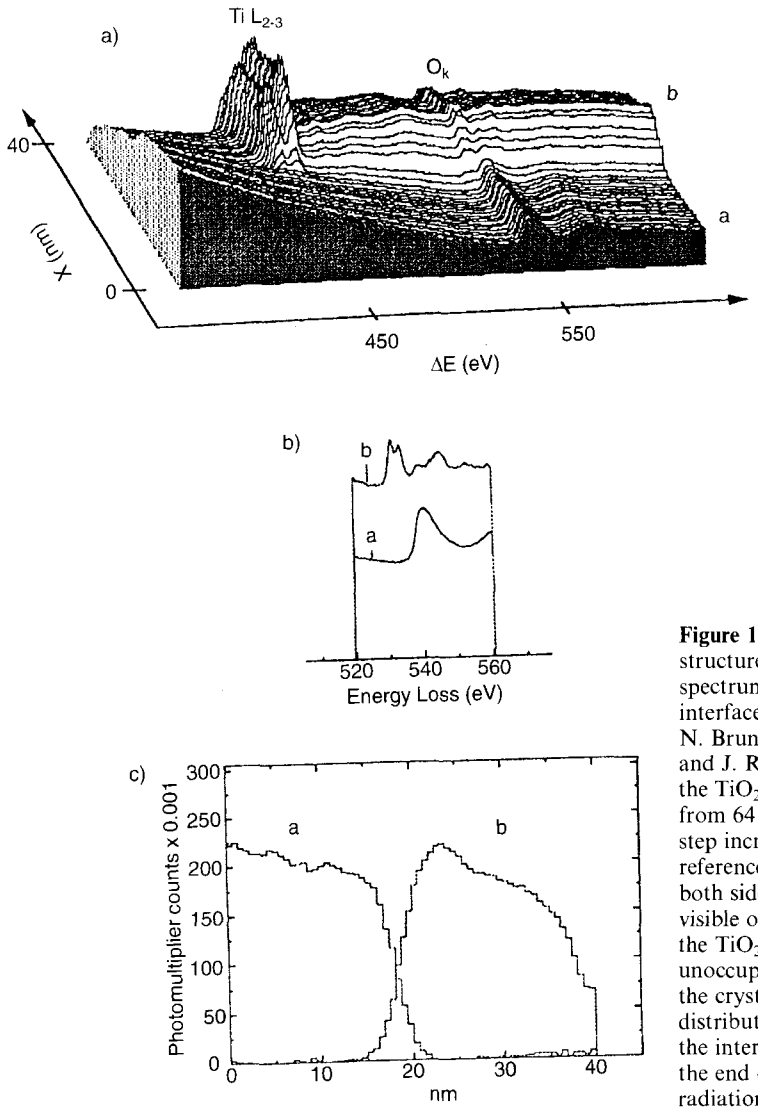


Figure 10. Mapping of the fine structures on a O K edge along a spectrum line across a SiO_2–TiO_2 interface (in collaboration with N. Brun, K. Yu-Zhang, D. Hurtaud and J. Rivory). (a) Spectrum line from the TiO_2 area into the SiO_2 area, made from 64 spectra (acquisition time 8 s, step increments 0.6 nm). (b) Selected reference O K edge profiles selected on both sides of the interface; the doublet visible on the front side of the edge in the TiO_2 case reflects the existence of unoccupied O p–Ti d-orbitals split by the crystal field effect. (c) MLS fit distribution of both references across the interface; the intensity decrease at the end of the scan is due to increased radiation damage on the TiO_2 side.

by selecting reference spectra in the sequence far from the regions where spectral variations are likely to occur). The definition can also be realized using pattern recognition and classification techniques which sort all the spectra of a sequence into families exhibiting a certain degree of similarity. One of these approaches, the use of artificial neural networks (ANNs) has recently been introduced for the analysis of near-edge structure components across $Si–SiO_2$ interfaces [60]. The basic idea is to consider any spectrum as a vector in a multidimensional space, the number of dimensions being equal to the number of energy-loss channels in the range of interest. Several criteria (e.g. scalar product and cross-correlation coefficient) can be used to select the basis vectors of this space, on which the projection coefficients of all spectra in the sequence provides a measurement of their weight in terms of the basis spectra. Other methods, such as the factorial analysis of correspondence, would provide other routes to classifying in a hierarchical order the variance of components within a large data set [61]. In all cases the full impact of these classification and analysis tools relies on a satisfactory physical understanding of the basis vectors or principal components.

It is obvious that these novel possibilities will open up quite large fields of, as yet undefined, applications. Interfaces, multilayers, nano-objects, and clusters constitute classes of objects particularly well suited to such subnanometer mapping of electronic structures. Among the many problems under present investigation [62] is the mapping of the p–d electron-state hybridization across a $SiO_2–TiO_2$ interface (Fig. 10).

The future of EELS mapping therefore relies on several developments in terms of instrumentation, with access to atomic-size electron probes such as the one demonstrated by Batson [63] and theoretical interpretation, applying for instance to the low-loss domain (maps of dielectric constants, optical coefficients, adhesion forces, etc.).

Acknowledgements

This review is the fruit of the collaboration of many colleagues over two decades, and I am very pleased to thank all of them who have contributed at different stages: R. Castaing, B. Jouffrey, P. Trebbia, C. Mory, C. Jeanguillaume, O. L. Krivanek, M. Tencé, N. Bonnet, E. Lefèvre, M. G. Walls, D. Bouchet, N. Brun, O. Stephan, and P. Redlich.

1.3.5 References

[1] H. Raether, *Springer Tracts Modern Phys.* **1965,** *38*, 85.
[2] S. E. Schnatterly, *Solid State Phys.* **1979,** *14*, 275.
[3] H. Raether, *Springer Tracts Modern Phys.* **1980,** *88*, 1.
[4] P. Schattschneider, *Fundamentals of Inelastic Electron Scattering*, Springer Verlag, Wien, **1986.**
[5] J. Fink, *Topics Appl. Phys.* **1992,** *69*, 203.
[6] J. Hillier, R. F. Baker, *J. Appl. Phys.* **1944,** *15*, 663.
[7] D. B. Wittry, *Br. J. Appl. Phys.* **1969,** *3*, 1757.
[8] A. V. Crewe, M. S. Isaacson, D. E. Johnson, *Rev. Sci. Instrum.* **1971,** *42*, 411.
[9] O. L. Krivanek, P. R. Swann, in *Quantitative Microanalysis with High Spatial Resolution* (Eds. G. W. Lorimer, M. H. Jacobs, P. Doig), The Metals Society, London **1981,** p. 136.
[10] O. L. Krivanek, C. C. Ahn, R. B. Keeney, *Ultramicroscopy* **1987,** *22*, 103.
[11] D. Bouchet, C. Colliex, P. Flora, O. L. Krivanek, C. Mory, M. Tencé, *Microsc. Microanal. Microstruct.* **1990,** *1*, 443.
[12] O. L. Krivanek, C. Mory, M. Tencé, C. Colliex, *Microsc. Microanal. Microstruct.* **1991,** *2*, 257.

[13] R. D. Leapman, in *Transmission Electron Energy Loss Spectrometry in Materials Science* (Eds. M. M. Disko, C. C. Ahn, B. Fultz), TMS, Warrendale **1992,** p. 47.
[14] M. K. Kundmann, O. L. Krivanek, *Microsc. Microanal. Microstruct.* **1991,** *2*, 245.
[15] P. E. Batson, *Rev. Sci. Instrum.* **1986,** *57*, 43.
[16] R. Castaing, L. Henry, *C. R. Acad. Sci., Ser. B* **1962,** *255*, 76.
[17] A. El Hili, *J. Microsc.* **1966,** *5*, 669.
[18] C. Colliex, B. Jouffrey, *C. R. Acad. Sci., Ser. B* **1970,** *270*, 673.
[19] S. Senoussi, L. Henry, R. Castaing, J. *Microsc.* **1971,** *11*, 19.
[20] G. Zanchi, J. P. Perez, J. Sévely, *Optik* **1975,** *43*, 495.
[21] H. Rose, W. Pejas, *Optik* **1979,** *54*, 235.
[22] S. Lanio, *Optik* **1986,** *73*, 99.
[23] H. Shuman, A. P. Somlyo, in *Analytical Electron Microscopy* (Ed. R. H. Geiss), San Francisco Press, San Francisco, CA **1981,** 202.
[24] O. L. Krivanek, A. J. Gubbens, N. Dellby, *Microsc. Microanal. Microstruct.* **1991,** *2*, 315.
[25] O. L. Krivanek, A. J. Gubbens, N. Dellby, C. E. Meyer, *Microsc. Microanal. Microstruct.* **1992,** *3*, 187.
[26] C. Jeanguillaume, C. Colliex, *Ultramicroscopy* **1989,** *28*, 252.
[27] J. A. Hunt, D. B. Williams, *Ultramicroscopy* **1991,** *38*, 47.
[28] R. D. Leapman, J. A. Hunt, in *Microscopy: The Key Research Tool*, EMSA **1992,** p. 39.
[29] R. D. Leapman, J. A. Hunt, R. A. Buchanan, S. B. Andrews, *Ultramicroscopy* **1993,** *49*, 225.
[30] J. L. Lavergne, J. M. Martin, N. Belin, *Microsc. Microanal. Microstruct.* **1992,** *3*, 517.
[31] C. Colliex, M. Tencé, E. Lefèvre, C. Mory, H. Gu, D. Bouchet, C. Jeanguillaume, *Mikrochim. Acta* **1994,** *114/115*, 71.
[32] M. Tencé, M. G. Walls, C. Jeanguillaume, C. Colliex, X. Thomas, O. Jbara, J. Cazaux, *Inst. Phys. Conf. Ser.* **1989,** *98*, 311.
[33] C. Colliex, *Adv. Opt. Electron Microsc.* **1984,** *9*, 65.
[34] R. F. Egerton, *Electron Energy Loss Spectroscopy in the Electron Microscope*, Plenum, New York, **1986.**
[35] M. M. Disko, C. C. Ahn, B. Fultz, *Transmission Electron Energy Loss Spectrometry in Materials Science*, TMS, Vol. 2, Warrendale, **1992.**
[36] R. F. Egerton, *Ultramicroscopy* **1979,** *4*, 169.
[37] R. D. Leapman, P. Rez, D. F. Mayers, *J. Chem. Phys.* **1980,** *72*, 1232.
[38] F. Hofer, *Microsc. Microanal. Microstruct.* **1991,** *2*, 215.
[39] H. Shuman, A. P. Somlyo, *Ultramicroscopy* **1987,** *21*, 23.
[40] R. D. Leapman, C. R. Swyt, *Ultramicroscopy* **1988,** *26*, 393.
[41] T. Manoubi, M. Tencé, M. G. Walls, C. Colliex, *Microsc. Microanal. Microstruct.* **1990,** *1*, 23.
[42] R. D. Leapman, D. E. Newbury, *Anal. Chem.* **1993,** *65*, 2409.
[43] L. Reimer, I. Fromm, C. Hülk, R. Rennekamp, *Microsc. Microanal. Microstruct.* **1992,** *3*, 141.
[44] C. Jeanguillaume, P. Trebbia, C. Colliex, *Ultramicroscopy* **1978,** *3*, 237.
[45] C. Jeanguillaume, M. Tencé, P. Trebbia, C. Colliex, *Scanning Electron Microscopy*, Vol. 2, SEM Inc, AMF O'Hare, Chicago, IL **1983,** p. 745.
[46] O. L. Krivanek, A. J. Gubbens, M. K. Kundmann, G. C. Carpenter, *Proc. 51st EMSA Meeting* (Eds. G. W. Bailey, C. L. Rieder), San Francisco Press, San Francisco, CA **1993,** p. 586.
[47] F. Hofer, P. Warblicher, W. Grogger, *Ultramicroscopy* **1995,** *59*, 15.
[48] P. A. Crozier, *Ultramicroscopy* **1995,** *58*, 157.
[49] N. Bonnet, C. Colliex, C. Mory, M. Tencé, *Scanning Microscopy Supplement 2*, Scanning Microscopy International, AMF O'Hare, Chicago, IL **1988,** p. 351.
[50] C. Mory, C. Colliex, *Ultramicroscopy* **1989,** *29*, 339.
[51] B. Bhanu Prasad, D. Jeulin, C. Daly, C. Mory, M. Tencé, C. Colliex, *Microsc. Microanal. Microstruct.* **1991,** *2*, 107.
[52] C. Colliex, *Microsc. Microanal. Microstruct.* **1991,** *2*, 403.
[53] A. Berger, J. Mayer, H. Kohl, *Ultramicroscopy* **1994,** *55*, 101.
[54] H. Kurata, S. Isoda, T. Kobayashi, *Microsc. Microanal. Microstruct.* **1995,** *6*, 405.
[55] M. Tencé, M. Quartuccio, C. Colliex, *Ultramicroscopy* **1995,** *58*, 42.
[56] R. D. Leapman, J. A. Hunt, R. A. Buchanan, S. B. Andrews, *Ultramicroscopy,* **1993,** *49*, 225.
[57] H. Müllejans, J. Bruley, *Ultramicroscopy* **1994,** *53*, 351.
[58] J. Bruley, M. Tseng, D. B. Williams, *Microsc. Microanal. Microstruct.* **1995,** *6*, 1.
[59] C. Scheu, G. Dehm, H. Müllejans, R. Brydson, M. Rühle, *Microsc. Microanal. Microstruct.* **1995,** *6*, 19.
[60] C. Gatts, G. Duscher, H. Müllejans, M. Rühle, *Ultramicroscopy* **1995,** *59*, 229.
[61] P. Trebbia, N. Bonnet, *Ultramicroscopy* **1990,** *34*, 165.
[62] N. Brun, **1995,** private communication.
[63] P. Batson, *Nature* **1993,** *366*, 727.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

1.4 High Voltage Electron Microscopy

1.4.1 Introduction

Because the behavior of materials is determined by lattice defects (point defects, dislocations, etc.), much attention has been paid to their direct observation, particularly to individual dislocations [1, 2]. Lattice defects, however, have strong strain fields associated with them in crystals, and thus their behavior is very sensitive to the thickness of the specimen. Figures 1 and 2 show the thickness effect on the dislocation density [3] and the cell formation [4] in aluminum crystals, respectively. The dislocation density decreases quite remarkably (Fig. 1) when the thickness becomes smaller than a critical value (i.e., 0.4–0.8 μm in cold-worked aluminum). This is a function of both the kind of material and the dislocation density [3]. With no applied stress, the position of each individual dislocation is determined by the local internal stress, which is a function of the dislocation density and the interaction among different families of dislocations in the local region, in addition to the image force. Therefore, to observe the same distribution of dislocations as that in bulk materials, the critical thickness increases more than that required for the dislocation density [5]. The formation of cell structures depends on the mutual interaction among moving dislocations, and thus the specimen thickness for this process must be larger than the mean free path of dislocations. Thus, the critical thickness increases further to a few micrometers to become representative for the cell structures of bulk specimens [5]. For example, the critical thickness in Fig. 2 is about 3 μm. Additionally, recovery and recrystallization, phase transformations, etc. are also very sensitive to the specimen thickness [3].

The above facts show that not only the behavior of lattice defects but also their density markedly changes when the specimen thickness is smaller than the critical value which is a function of the behavior of related lattice defects.

However, the maximum observable specimen thickness with the 100 kV class electron microscopes (EMs) is smaller than these critical thicknesses in general, especially for the dislocation behavior [5]. For this reason there is a need for high voltage electron microscopes (HVEMs) for which the maximum observable thickness of the specimens is larger than the critical values. For such purposes, a few 0.5 MV HVEMs were constructed despite the technical difficulties (which are considered to be roughly proportional to the third power of the accelerating voltage required to keep the same mechanical and electrical stability at higher accelerating voltages).

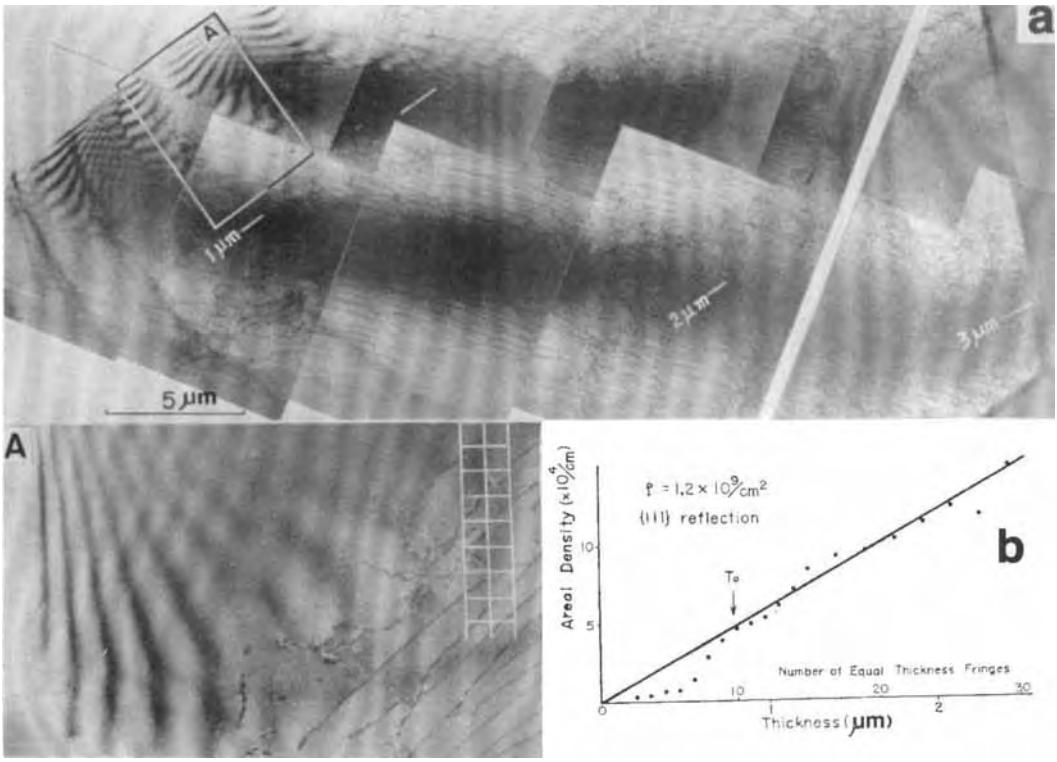


Figure 1. Thickness effect on the dislocation density in aluminum [3]. (a) Specimen was finished wedge-shaped after deformation at room temperature; micrograph was taken at 0.5 MV. Equal-thickness fringes appear clearly near the specimen edge at the left-hand side. Micrograph A shows an enlargement of the framed part A; (b) shows the relationship between the areal density of dislocations and the specimen thicknesses indicated in micrograph (a).

In 1965 it was found that the same dislocation behavior as that in bulk materials can be observed at 0.5 MV in light metals whose atomic number is smaller than about 20, and in situ experiments were carried out on various phenomena in such metals [3, 5].

Since 1965, the HVEMs and their related accessaries have rapidly developed and widely applied to the natural sciences, especially materials science. Two ultra-HVEMs operating at 3 MV have been installed independently in Osaka and Toulouse in 1970, and, to date, about 60 HVEMs, whose accelerating voltages are higher than 0.5 MV, have been installed around the world [6].

This account deals with the usefulness and the applications of HVEMs, mainly in the field of materials science.

1.4.2 Voltage Dependence of Operational Features of Electron Microscopes

It was shown experimentally [7] that with increasing accelerating voltage both

Figure 2. Thickness effect on the formation of cell structures in aluminum [4]. The wedge-shaped specimen was deformed by HVEM to form cell structures, and the micrograph was taken at 0.5 MV. Local thicknesses of the specimen are indicated in the micrograph.

spherical and chromatic aberrations decrease remarkably as well as the wavelength of electrons. Furthermore, the following advantages can be obtained at higher accelerating voltages: (a) a decrease in contamination of the specimens during observation [8]; (b) a decrease in irradiation damage in biological and polymer specimens [9]; (c) a decrease in temperature rise of the specimens during observation [6]; and (d) an increase in the simultaneous excitation of many waves. Point (d) is closely related to a phenomenon called electron channeling [10] or the Bloch wave channeling [11, 12], as mentioned below. Increase of the maximum observable thickness and visibility of lattice defects in adjacent different crystals at the same time are also related to this phenomenon [6]. Additional advantages are: (e) an easy determination of the structure factor of crystals [13, 14] and (f) accurate studies of electron irradiation damage in materials [15, 16]. With regard to (f) electron irradiation damage during observation can be effectively suppressed under suitable conditions, as mentioned below [6]. There are a number of other applications [6].

1.4.2.1 Electron Channeling at High Accelerating Voltages

A theoretical estimation was made in 1962 of the voltage dependence on the penetration power of electrons [17]. The result showed that the penetration power is almost saturated at about 500 kV, and that it increases only 3.3 times as compared to that at 100 kV even when the accelerating voltage increases up to infinity. Many electron microscopists concluded from these results that the penetration power is proportional to the maximum observable thickness of the specimens, and this resulted in a pessimistic estimation of the voltage dependence on the maximum observable thickness.

However, many-beam effects become very important for the maximum observable thickness at high voltages. When the

accelerating voltage increases, a bright band appears at the symmetry position of bend contours. This phenomenon is due to the many-beam effect on the anomalous transmission of electrons. In the case of aluminum, the anomalous transmission at the symmetry position of (1 1 1) systematic reflections begins to appear clearly at about 1 MV, and its intensity remarkably increases with increasing accelerating voltage [6]. Figure 3 shows the bright-field image (BF) and a dark-field image (DF) of the (1 1 1) extinction contour at 3 MV.

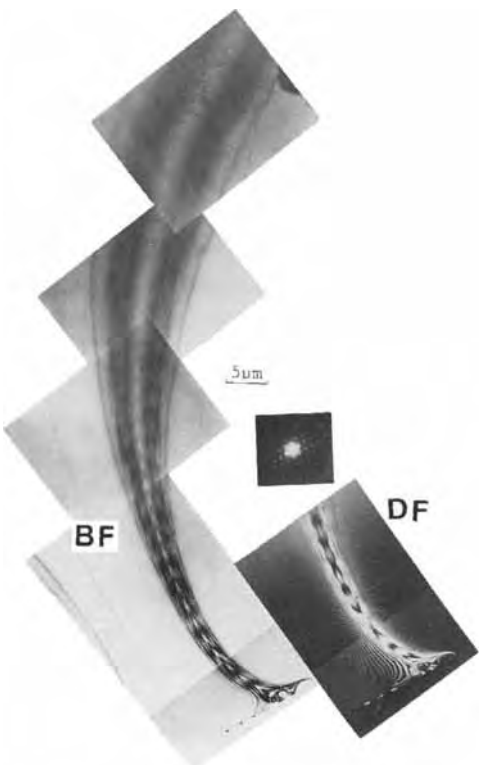


Figure 3. Anomalous transmission of electrons at the symmetry position of (1 1 1) reflections in aluminum at 3 MV [6]. Micrographs at the left- and right-hand sides show the bright (BF) and dark-field (DF) images, respectively. An electron diffraction pattern was taken at the symmetry position.

A bright band appears at the symmetry position of (1 1 1) systematic reflections in the BF, and at both sides of the (1 1 1) contour in the DF.

The axial and planar channeling of electrons can be estimated from the effective width of Kikuchi bands [12]. The axial (α_a) and planar (α_p) channeling parameters are defined as follows:

$$\alpha_a = \frac{4mZ_1Z_2e^2}{h^2Nd^2} \tag{1}$$

$$\alpha_p = \frac{4mV_h}{h^2b_h^2} \tag{2}$$

where $Z_1 = 1$, and m, Z_2, e, h, N and d in Eq. (1) are the relativistic mass of an electron, the atomic number of crystal atoms, the charge of an electron, Planck's constant, the atomic density, and the distance along the crystal axis, respectively, and V_h and b_h in Eq. (2) are the Fourier coefficient of the crystal potential and the reciprocal vector, respectively. Experimental results show that the classical model is adequate when these two parameters are larger than unity (i.e., $\alpha_p > 1.1$) and that the degree of anomalous transmission at the symmetry position of reflections is also well correlated with the value of α_p irrespective of both the kind of crystal and the indices of the reflection [18].

Table 1 shows the calculated α_a and α_p by Eqs. (1) and (2). The values of both channeling parameters are proportional to the electron intensity passing through the material. Thus it is concluded, from Eqs. (1) and (2), that the electron intensity increases linearly with increasing accelerating voltage, because the relativistic mass of electrons linearly increases with the accelerating voltage. This means that the maximum observable specimen thickness

Table 1. (a) Calculated axial channeling parameter, α_a.

Voltage (MV)	Beam direction		
	⟨110⟩	⟨200⟩	⟨222⟩
0.5	3.1_7	1.5_8	0.5_2
1.0	4.7_4	2.3_7	0.7_9
1.5	6.3_1	3.1_5	1.0_5
2.0	7.8_8	3.9_4	1.3_1
2.5	9.4_5	4.7_2	1.5_7
3.0	11.0_2	5.5_1	1.8_3

Table 1. (b) Calculated planar channeling parameter, α_p

Voltage (MV)	Order of reflection		
	{111}	{200}	{220}
0.5	0.8_9	0.5_5	0.1_7
1.0	1.3_3	0.8_2	0.2_6
1.5	1.7_7	1.1_0	0.3_5
2.0	2.2_1	1.3_7	0.4_3
2.5	2.6_5	1.6_4	0.5_2
3.0	3.0_9	1.9_2	0.6_1

is expected to increase linearly with increasing accelerating voltage when only the electron intensity is taken into account.

The electron intensity also increases with increasing the atomic density along the electron path, that is, with decreasing d in Eq. (1) and b_h in Eq. (2), respectively.

Two important phenomena, due to electron channeling, are: (a) the orientation dependence of electron irradiation damage and (b) the straight path of electrons along the incident direction in materials. According to (a), the rate of electron irradiation damage is proportional to the electron intensity, and thus the damage rate remarkably decreases when higher order reflections and/or higher order directions are used, as mentioned above [19]. Furthermore, an image-intensifier system is always used for the observation of images with the HVEMs. Thus, the intensity of the incident beam can be decreased by more than three orders of magnitude. Specimen damage during observation can therefore be suppressed to an extremely small level. According to (b), electron channeling occurs such as the total reflection on inner surfaces of a fine column on an atomic scale along the incident electron beam, as recognized by the electron irradiation induced foreign-atom implantation (below).

1.4.2.2 The Objective Aperture Effect on the Maximum Observable Specimen Thickness

As already mentioned, electron intensity, transmitted through materials by electron channeling, increases linearly with increasing accelerating voltage. Practically, however, various aberrations of electron microscope images must be taken into account for the determination of the maximum observable specimen thickness. Here, the maximum observable thickness is discussed under a fixed image resolution.

The total aberration (δ_T) of electron microscope images consists of the diffraction (δ_{diff}), spherical (δ_{sph}) and chromatic (δ_{chr}) aberrations. This is expressed as

$$\delta_T = (\delta_{diff}^2 + \delta_{sph}^2 + \delta_{chr}^2)^{1/2} \qquad (3)$$

where $\delta_{diff} \sim 0.6\lambda/\alpha$, $\delta_{sph} = C_s\alpha^3$, and $\delta_{chr} = C_c\alpha\,\Delta E/E$. Here, λ, α, C_s and C_c, and E and ΔE are the electron wavelength, effective aperture angle, spherical and chromatic aberration coefficients, and energy and energy fluctuation of electrons, respectively. In these three terms, δ_{diff} is divided by α, but the terms δ_{sph} and δ_{chr} are

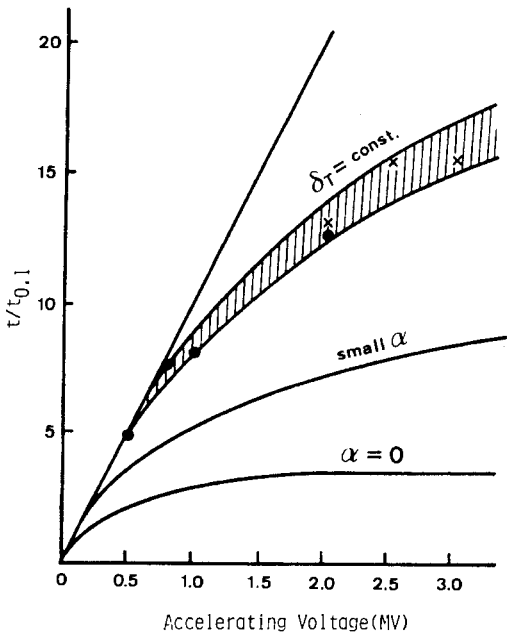


Figure 4. Voltage dependence of the maximum observable thickness of specimens [6, 20]. The maximum observable thickness at each voltage (t) is normalized against that at 100 kV ($t_{0.1}$). The hatched curve corresponds to measurements at a fixed resolution. Marks in the hatched curve show measured values on various metals and alloys.

multiplied by α^3 and α, respectively. Thus, there is an optimum aperture size that minimizes δ_T given by Eq. (3), depending on the accelerating voltage. Under this condition, the diffraction contrast of images clearly results from higher order excited waves, which are cut off by the objective aperture, because their intensity markedly increases at high voltages. Figure 4 shows the aperture effect on the image contrast as a function of the accelerating voltage [20]. The hatched band shown in Fig. 4 was obtained experimentally for the dislocation images in various materials with a fixed image resolution of 0.5 nm [6]. The lowest curve of $\alpha = 0$ in Fig. 4 corresponds to a case in which only one wave is excited, that is, it represents the voltage dependence of the penetration power of electrons.

It is noted in Fig. 4 that the maximum observable thickness of specimens at 3 MV practically increases more than 15 times compared to that at 100 kV. In other words, $t/t_{0.1} > 15$, even when the image resolution is fixed, while it deviates from a linear relation for accelerating voltages higher than about 0.5 MV [6]. The value at 3 MV corresponds to about 3 μm in thickness even in high Z materials such as tungsten and gold. The same behavior of all lattice defects as that in bulk materials can be observed in situ with 3 MV class HVEMs, even when the atomic number of the constitutive atoms is extremely high.

1.4.3 In Situ Experiments with High Voltage Electron Microscopes

1.4.3.1 Specimen Treatment Devices for In Situ Experiments

As previously stated, phenomena representative for bulk materials can be investigated dynamically with HVEMs. Therefore, to carry out in situ experiments, various specimen treatment devices have been improved [6, 21]. At present, specimen manipulations such as stretching, alternating stressing, constant-load deformation, electron irradiation, treatments in various environments, and their combinations can be carried out in a wide temperature range from about 5 to 2300 K. All of these devices are of top-entry type and

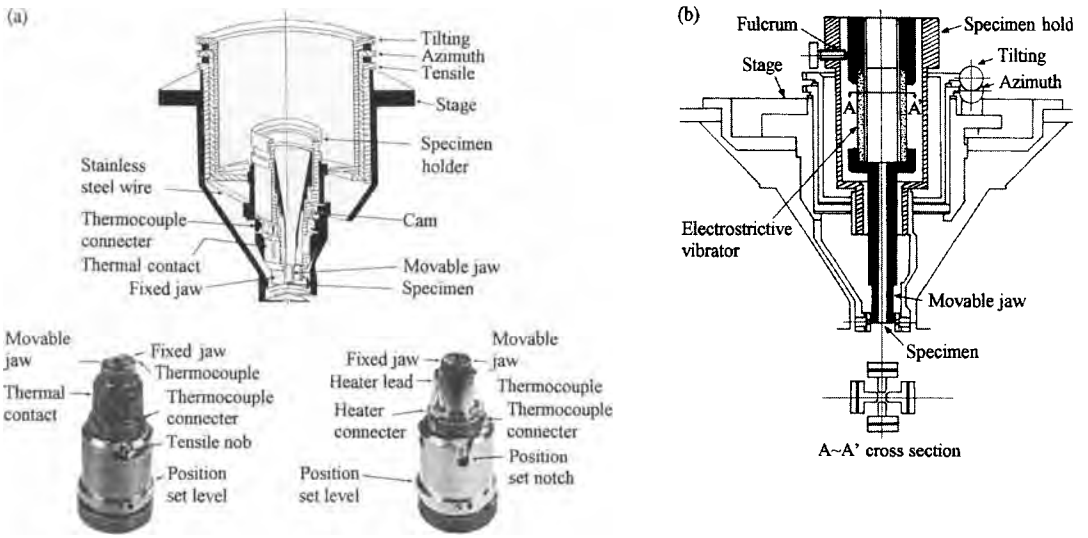


Figure 5. In situ deformation devices [6]. (a) Stretching devices for deformation at high (right) and low (left) temperatures, respectively. These devices are placed on a goniometer stage. (b) Cross-sectional view of an alternating stressing device using an electrostrictive vibrator.

mounted on a universal goniometer stage whose maximum tilting angle is 10° or more [6]. Furthermore, most of these specimen holders can be easily exchanged using an air-lock system.

Figures 5–8 are some examples of those specimen treatment devices for in situ experiments [6]. Since the objective pole piece of the HVEMs is concave to decrease the chromatic and spherical aberrations, the path necessary for the specimen transfer is generally very large. However, the specimen chamber is so large that the various devices mentioned above can easily be mounted on the specimen stage, also a sufficient thermal and electromagnetic shielding as well as a long movement of the specimen holder jaws becomes possible.

Figures 5a and 5b show deformation devices for stretching and alternating stressing in fatigue deformation, respectively. Most of these devices can deform the specimens up to about 1 kg in loading in a temperature range from about liquid nitrogen temperature to about 1200 K. Figure 6 shows a sectional diagram of a deformation stage for use at very high temperatures. The specimen is set on the tip of a tantalum cylindrical cone, and power for heating is supplied to the tantalum cone by electron beam heating with a tungsten ring-filament. Using this stage, the specimen can be tilted around any axis up to 8° during specimen treatment, and deformed to 50% elongation in a temperature range from room temperature to 2300 K. This stage has been effectively used for studying high temperature characteristics of refractory materials including various ceramics and their composites, as will be mentioned in a later section.

Figure 7 shows a sectional diagram and an outside view of a liquid helium temperature stage, on which the specimen can be tilted up to 8° around any axis during

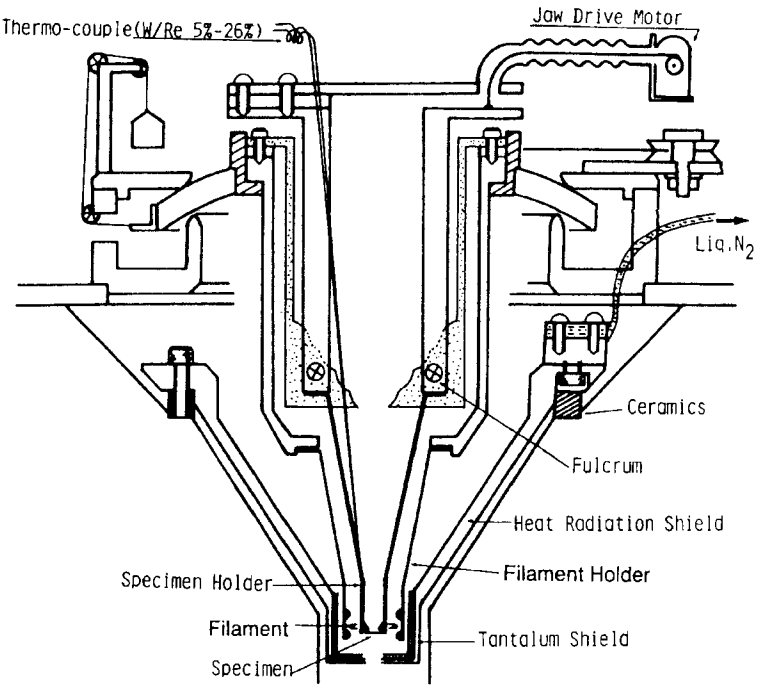


Figure 6. A very high temperature deformation device [6]. The specimen can be stretched up to 50% strain and tilted around any axis up to 8° in a temperature range from room temperature to about 2300 K.

the experiment. The stage was improved further to allow the specimens to stretch up to about 50% strain in a temperature range from room temperature to about 10 K.

Figure 8a shows a sectional diagram of a universal environmental cell that can be used for both biological and physical specimens [6]. An enlargement of the main part of the cell that consists of three films x, y, and z can be seen in Fig. 8b. The cell films are vapor-deposited aluminum, silicon oxide, and carbon films, depending on the purpose, and they are supported by metal meshes (300 and 400 mesh in size). Gases or liquids are circulated through the capsule by two pipes and the capsule is pressurized up to 300 kPa. The specimen holder can be exchanged by an air-lock system. In Fig. 8b, the specimen is set either directly on the center nickel mesh (y) or on a vapor-deposited film coated onto the nickel mesh. The specimen is heated by directly passing an electric current through the nickel mesh, and cooled by connecting the mesh with a liquid nitrogen tank. The specimen also can be heated to high temperatures (>1300 K) by passing a current directly through the specimen. Films x, y, and z are separated 30–50 μm from each other to allow gas exchange between these gaps. Furthermore, the specimen drift is less than 0.1 nm s^{-1} at room temperature. The environmental cell [6] has been effectively used for studying: (a) the sintering process of fine metal powders which are sensitive to oxidation and reduction; (b) chemical reactions, such as oxidation and inner oxidation, among gases, liquids, and solids; (c) the determination of the atomistic structure of liquids; (d) the observation of wet biological and microbiological specimens and their movement in salt water.

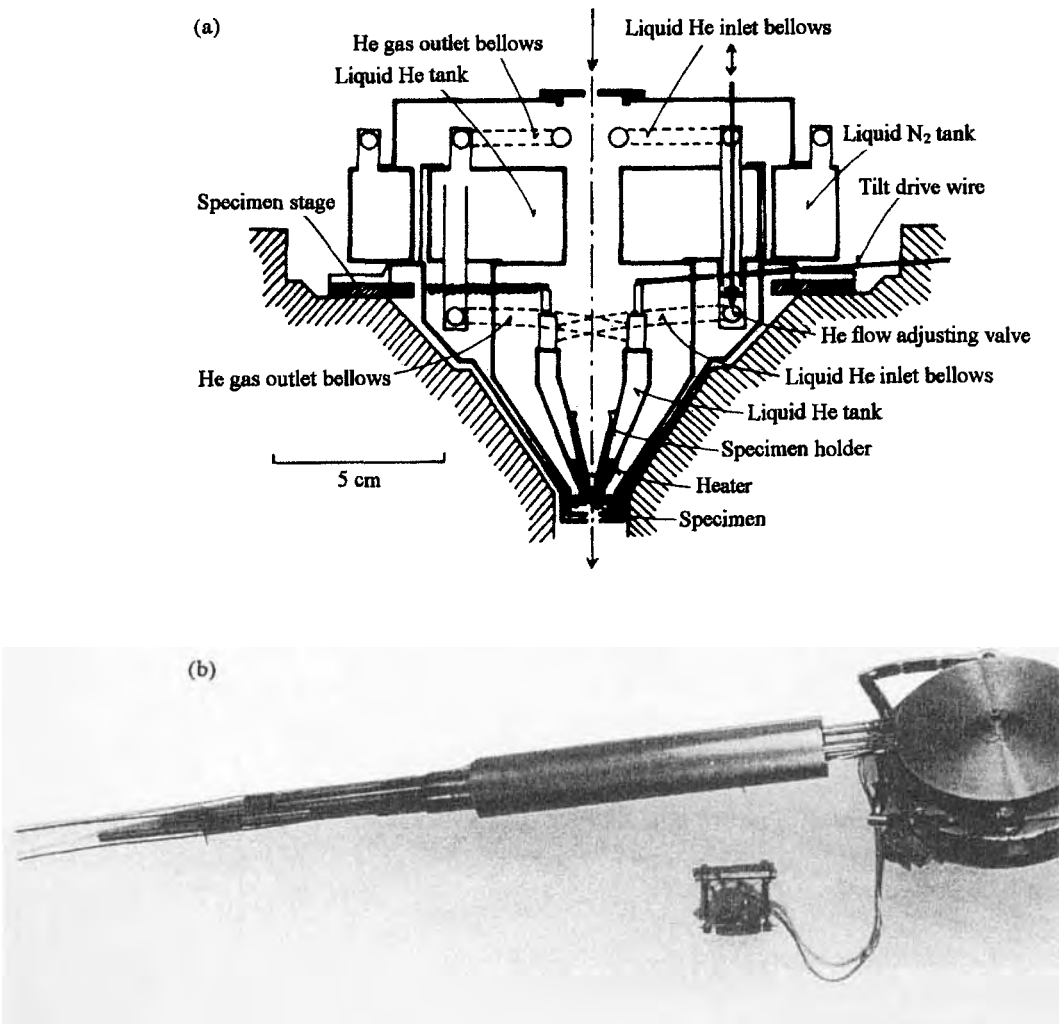


Figure 7. A sectional diagram of a liquid helium temperature stage [6]. The specimen can be tilted up to 8° around any axis during cooling.

1.4.3.2 Applications to Materials Science

As stated, to carry out in situ experiments the specimen thickness must be larger than the mean free path of related lattice defects, otherwise doubtful conclusions will result. Practically, at least 1.5 MV class HVEMs are necessary for studying bulk representative dislocation behavior in materials whose atomic number is larger than about 25 (i.e., most of industrially important metals) [6, 31]. In the case of 3 MV HVEMs, the maximum observable thickness is always larger than the critical thickness for various phenomena even in very high Z materials.

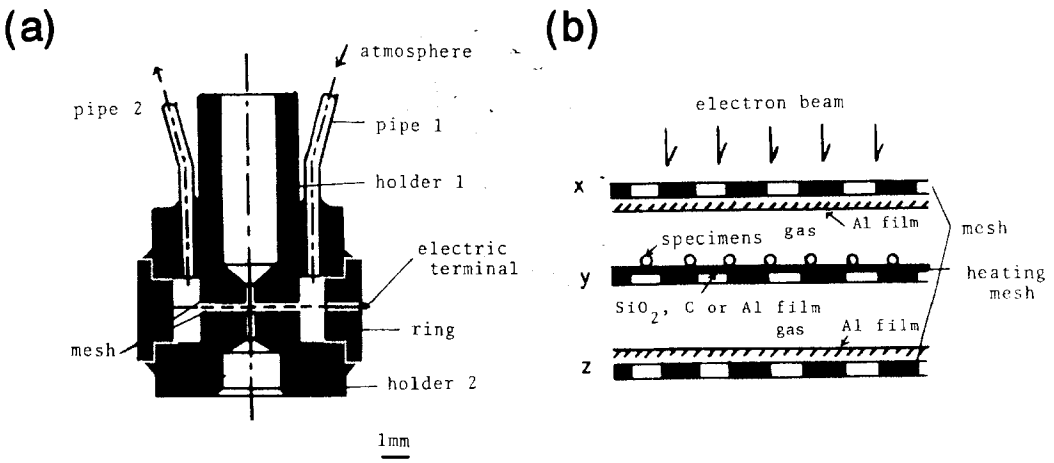


Figure 8. (a) Sectional diagram of a universal environmental cell [6]. (b) Enlargement of a main part of the cell, which contains the specimen. This part consists of three films x, y and z. Maximum atmospheric pressure 300 kPa; maximum specimen temperature 1300 K.

Deformation [6, 21–31]

In situ deformation studies give valuable information on the following phenomena: (a) the behavior of individual dislocations detected by their slip traces; (b) the shape of dislocations under stress from which the magnitude of the local stress can be estimated; and (c) dislocation interactions and interaction between dislocations and microstructures. Based on these studies, new information has been obtained on the mechanisms of the various phenomena listed below.

(1) *Uniaxial Deformation*

At the onset of deformation, long dislocation segments of many slip systems are activated irrespective of their Schmid's factors immediately before the apparent yield stress. When the applied stress approaches the apparent yield stress the number of activated dislocation sources increases. In this process, the following facts become evident: (a) The source dislocation cross-slips so frequently that the emitted dislocations have many super-jogs, as seen in Fig. 9 [6]. (b) Since the shear stress necessary to move the emitted dislocations rapidly increases as a result of (a), the pole dislocations are also moved outwards when the stress slightly increases, even when their Schmid factors are very small. Thus, only several dislocations are multiplied from a dislocation source in metals with a relatively high stacking fault energy [6]. (c) Aside from grown-in dislocation sources, dislocation sources are newly formed not only by double cross-slip of moving dislocations but also by the interaction between dislocations of different families. They are also formed by the interaction between a mobile dislocation and other microstructures such as small precipitates. These newly formed dislocation sources increase in number with increasing stress. Since the emitted dislocations from the sources have a large number of super-jogs in general, the passage through forest dislocations

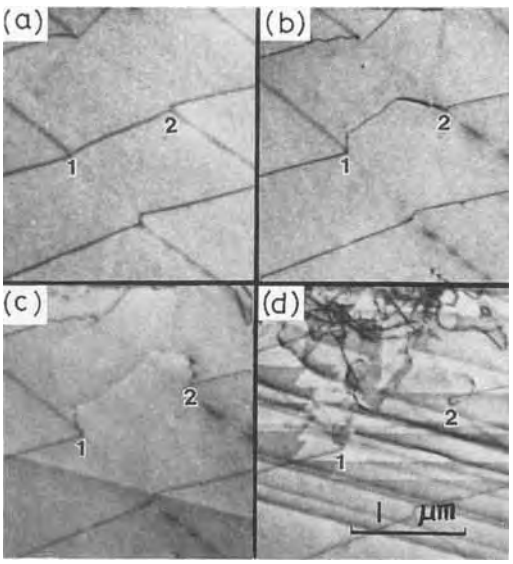


Figure 9. Activation of a grown-in dislocation source due to cross slip in aluminum [6]. The micrograph was taken at 0.5 MV. When the dislocation source (1–2) begins to move, the dislocation becomes zigzag in shape by cross-slip in (b). Exact shape of the activated dislocation source can be determined from a zigzag slip trace in (c). (d) Slip traces of the dislocations emitted from single ended dislocation sources 1 and 2.

does not present a strong resistance against mobile dislocations until the spacing among forest dislocations becomes smaller than that of the superjogs. (d) The formation mechanisms of various dislocation structures in stages I, II and III were also verified in detail, and the work hardening mechanism was investigated as a function of the type of crystals, the specimen orientation and the stacking fault energy. In these results, the role of conjugate slip in the work hardening process was verified using crystals with very low stacking fault energy [6]. Namely, the conjugate slip pays an important role in the work hardening in stage II, but it contributes to the dynamic recovery in stage III.

(2) *Deformation Under Alternating Stress*
The behavior of individual dislocations is also revealed in this process, and it was found that the irreversible motion of screw dislocation plays an important role in fatigue deformation. In addition, the formation process and the dislocation structures of ladder structures were investigated in detail. These results provided useful information about the mechanism of fatigue fracture in materials [6].

(3) *Various Influences on Deformation*
Deformation is very sensitive not only to the specimen orientation but also to both deformation temperature and impurity atoms. A typical example of temperature dependence is the deformation of ⟨1 0 0⟩ and ⟨1 1 0⟩ oriented b.c.c. crystals, as shown in Fig. 10 [6]. The upper figures in Fig. 10 show the stress-strain curves taken at temperatures of $>0.2\,T_m$ and $<0.1\,T_m$ (where T_m is the absolute temperature of the melting point), respectively. The micrographs in Fig. 10b were taken from VTR films showing dislocation movement in the ⟨1 0 0⟩ oriented tungsten (upper pair) and iron (lower pair) crystals, respectively. From these micrographs it is concluded that the work hardening rate markedly decreases when only one slip system is operated in each individual local region of both crystals. That is, even in the ⟨1 0 0⟩ crystal, in which four slip systems are expected to be activated geometrically, only one slip system is operative at temperatures higher than $0.2\,T_m$, as seen at the right hand side of the micrographs in Fig. 10b. Furthermore, the temperature dependence of serrated yielding (the so-called Portevin–Le Chatelier effect) in various alloys has been investigated [6].

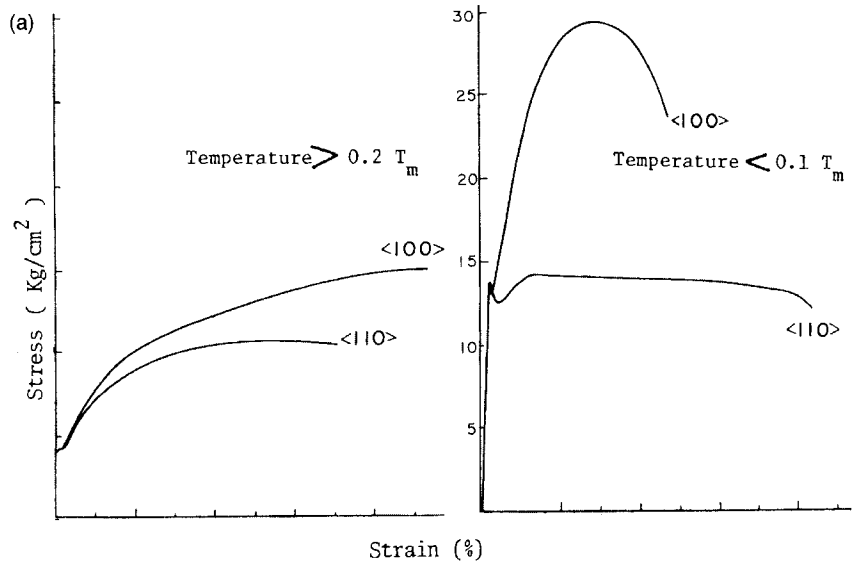


Figure 10. Temperature dependence of the work-hardening rate in the ⟨1 0 0⟩ and ⟨1 1 0⟩ b.c.c. crystals [6] (a). Micrographs were picked up from VTR films. Upper and lower series of micrographs show dislocation structures in the ⟨1 0 0⟩ oriented tungsten and iron crystals taken at the same homologous temperatures, respectively (b).

(4) *Other Processes*

Processes of the following phenomena have also been investigated in detail: (a) dislocation behavior in the vicinity of boundaries, (b) precipitation hardening, (c) deformation processes in various ceramics, (d) electron irradiation softening of b.c.c. metals due to enhancement of both multiplication and movement of dislocations [6], (e) the effect of hydrogen gas on the dislocation behavior in b.c.c. metals, (f) dislocation channeling in particle irradiated specimens, and so on. Furthermore, it was shown that so-called 'super plasticity' easily and rapidly (with a strain rate of about 10^{-2}) occurs when the grain diameter is kept in the range of 0.1 μm.

Annealing Phenomena [6, 19, 21–31]

These phenomena are generally determined by the diffusion of both point defects and solute atoms. When dislocations are involved in the phenomena as in the case of recovery and recrystallization, the specimen thickness must be large enough to keep the same distribution of dislocations as that in bulk specimens (i.e., a few μm) [6]. The critical thickness, however, remarkably decreases at low temperatures when the phenomena are only correlated with diffusion of the point defects and solute atoms, as in the case of Guinier–Preston (G.P.) zone formation, because of a remarkable decrease of the mean free path of these defects.

(1) *Recovery and Recrystallization*

Various deformed specimens were annealed in situ, and the rearrangement of dislocations, subgrain coalescence and the nucleation of new grains have been investigated in detail [6]. The results show that the misfit angle of boundary, which determines the boundary energy, increases with decreasing the dislocation density, which is directly related to the stored energy, within a recovered region during the nucleation process [32].

(2) *Sintering Process*

This phenomenon is very sensitive to the size of metal and ceramic powders. The coalescence among powders plays an important role when the powder size is ~0.1 μm in diameter or less [6]. Sintering and successive grain growth of ceramic composites have been also investigated in detail [19]. In ceramic composites whose melting point is higher than 2000 K, their phase diagrams can be determined by in situ annealing.

(3) *Precipitation Process*

The formation processes of G.P. zones [33] and precipitates have been investigated in situ in various metals and alloys, and useful information has been obtained for their formation mechanisms.

(4) *Melting Process*

The dislocation behavior during melting, and a peculiar phenomenon at solid–liquid interfaces [34] are important topics in the process.

Martensitic Transformations [19, 21–31, 35]

Phenomena related to this type of transformation are divided into two

groups, spontaneous and strain induced. Their formation mechanisms have been extensively investigated in both metals and ceramics by in situ experiments, especially by in situ deformation at low temperatures. From this new mechanisms of f.c.c. → h.c.p. and f.c.c. → b.c.c. transformations have been proposed. Furthermore, very toughened ceramics showing both plastic deformation and superplasticity at room temperature have been investigated from a view point of martensitic transformation [19].

Radiation Effects [6, 21–31]

The study of these phenomena is one of the important applications of in situ experiments, and the behavior of the primary and secondary defects induced by irradiation with high energy electrons and ions have been investigated extensively. This phenomenon is also correlated with both irradiation softening [6] of b.c.c. crystals and irradiation hardening of metals and alloys. Furthermore, the formation of nonequilibrium phases has also been carried out by the radiation effects, as mentioned later.

Environment-Material Interactions [6, 21–31]

The following subjects have been investigated by using both open- and closed-type environmental cells: (a) inner oxidation in metals and alloys, (b) hydrogen embrittlement of iron crystals, (c) chemical amorphization of metals [6], (d) crystallization of materials with low melting points and crystal growth of various oxides, (e) observation of wet biological and living microbiological specimens [6], and others. Closed-type environmental cells have been effectively used in this research field, because the pressure of atmospheres can be widely changed up to 300 kPa so that liquid materials can also be used [6].

Magnetic Domain Measurement [31]

Various sorts of magnetic domain structures and interaction between domain walls and microstructures have been investigated by changing the strength of the magnetic field during in situ observation.

Quantitative Studies [21–31, 36]

By using three-dimensional measurements, the distribution of microstructures such as complex dislocation structures, radiation induced secondary defects, small precipitates, and inclusions have been quantitatively determined. Furthermore, the strength of applied and internal stresses in local regions can be estimated from both the change in the radius of curvature and the movement of dislocations when the applied stress is changed. The dislocation velocity can also be measured as a function of the Burgers vector and the applied stress [36]. Furthermore, atomistic structures of very fine precipitates can also be accurately estimated after sufficient growth by in situ annealing.

1.4.4 New Research Fields Using High Voltage Electron Microscopy

The technical difficulty of keeping the same stability of electron energy at high accelerating voltages must be resolved in the construction of HVEMs. As a result, the functional features of conventional electron microscopes have been very much improved. In particular, high resolution electron microscopes operating at high voltages from 300 kV to 1.5 MV, compact HVEMs, high vacuum electron microscopes, high resolution analytical microscopes have been developed. Using a 300 kV electron microscope, high quality many-beam lattice images of diamond and silicon crystals have been obtained [37]. Furthermore, very stable high energy electrons can also be used in other studies, such as the formation of various nonequilibrium phases, various kinds of microfabrication, and the generation of synchrotron orbital radiation induced by the rosette motion of electrons around the nucleus of each atom.

One of the most important applications is the formation of non-equilibrium phases to cause new properties. Such nonequilibrium phases are synthesized by changing (a) the atomic structures and (b) the chemical composition. Extreme examples of (a) and (b) are the formation of amorphous solids and implantation of foreign atoms into various solids. Using the high energy electron irradiation induced (EII) method, various conditions for the formation of nonequilibrium solid phases can be easily and precisely controlled. Furthermore, compared with ion implantation and ion and neutron mixing methods, this method has the following advantages: (a) small irradiation damage, (b) small temperature rise (a few degrees) in the specimens, (c) deep implantation (more than 1 mm) of foreign atoms, and (d) very fine-scale implantation (i.e., about 1 nm in diameter). Furthermore, the foreign atom implantation can be carried out at any location within the material when the electron energy is sufficient for the implantation.

The formation mechanisms of these nonequilibrium phases can be dynamically investigated on an atomic scale by in situ experiments using sufficiently thick specimens. Today, modern electron microscopy, with the aid of HVEMs, forms a powerful tool for the characterization and identification of materials as well as finding an indispensable place in the 'microlaboratory', in which various sorts of specimen treatments, including the formation of nonequilibrium phases, can also be carried out precisely on an atomic scale.

1.4.4.1 Crystalline–Amorphous Transition

Figure 11 shows a crystalline–amorphous transition in the aluminum–zirconium system [32, 38]. Open circles represent the phases which have shown solid amorphization by the EII-method, and LQ indicates the composition range in which amorphization occurs by the liquid quenching method. From this, it can be easily recognized that the EII-method is superior to the liquid-quenching method in controlling the conditions of amorphization. The EII-method has been applied to examine the possibility of solid

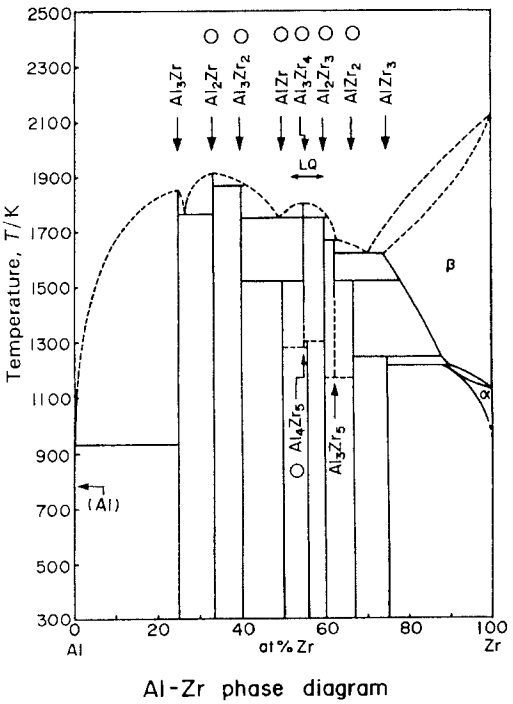


Figure 11. Crystalline–amorphous transition in the Al-Zr system [38]. Open circles represent the phases that have shown solid amorphization by EII. LQ indicates the composition range in which amorphization by the liquid quenching method takes place.

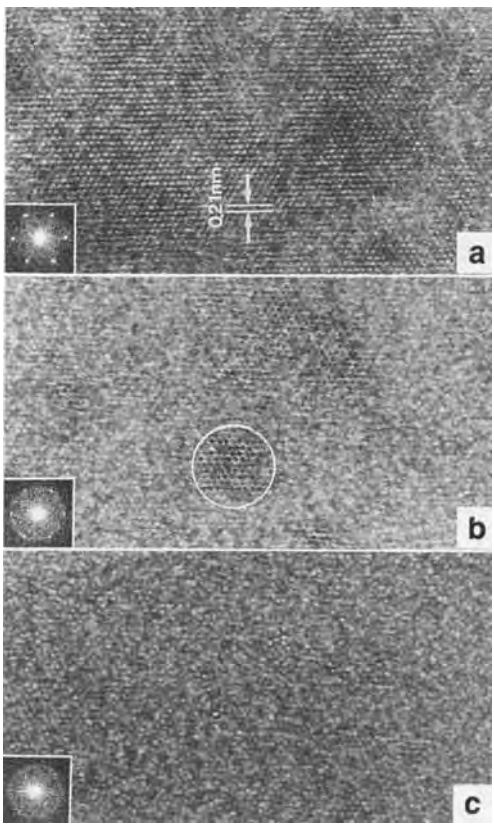


Figure 12. High resolution electron micrographs corresponding to start of solid amorphization (a). (b) Partly amorphitized. (c) completely amorphitized [32].

amorphization for more than 60 intermetallic compounds and about 30 ceramics. Figure 12 shows an example of the amorphization induced by electron irradiation at 2 MV in an Ni-50 at % Ti shape-memory alloy [32], which is very difficult to amorphize by quenching from the liquid phase. The bright spots in (a) and dark ones in (b) represent atom chains whose direction is parallel to the beam direction. It should be noted in Fig. 12b that periodic arrays of spots can be observed in very limited regions. The existence of a critical size for such a region can be concluded, because the periodic arrays of atom chains are only observed in blocks larger than a critical size. The critical size depends on the bonding mode and bond strength of the constitutive atoms (i.e., it decreases with increasing degree of covalent bonding and with decreasing coordination number of the atom) [32]. The critical size for the Ni-50 at % Ti alloy is about 3 nm in diameter, as shown by the white circle in Fig. 12b. Based on these experimental results, various factors determining amorphization have been clarified, and general rules for amorphization have been proposed [32, 39]. The general rules can be applied to all methods of amorphization such as chemical amorphization, liquid

quenching, vapor deposition, mechanical alloying, sputtering, particle irradiation and so on.

1.4.4.2 Electron-Irradiation-Induced Foreign Atom Implantation

The electron irradiation induced (EII) method has been used to implant foreign atoms into solids. A combination of solute and solvent atoms can be selected depending on the purpose, such as for the formation of supersaturated solid solutions, atom clusters and amorphous solids. For example, Si, Ge, Sn, and Pb atoms were implanted into an Al-matrix by the EII-method using 2 MV electrons at 160 K [32]. In these cases, amorphous phases are formed by mixing both Si and Ge atoms into the Al matrix. Small clusters of implanted atoms are also formed by slight annealing after the electron irradiation. Furthermore, it is concluded that the implantation speed markedly increases when amorphization occurs by mixing foreign atoms, as in the case of Al-Si and Al-Ge alloys.

Figures 13a and 13b were taken before and after implantation of Pb atoms into an Al crystal at 175 K, and Fig. 13c after post-annealing for 36 ks at 573 K [40]. It can be noted in Fig. 13b that the diffraction contrast does not appear so clearly in spite of the large difference of more than 20% in atom size between Pb and Al atoms. These implanted Pb atoms precipitate as ultra-fine particles by post-annealing, as shown in Fig. 13c.

When the foreign atom implantation by EII is carried out at slightly higher temperatures (e.g., at 293 K in Al-Pb alloys), the implantation of solute atoms hardly occurs, even after a large dosage of electrons [40]. It is therefore concluded that the implantation results from repeated short displacement of foreign atoms, and that the mobility of solute atoms is closely related to the behavior of point defects induced in the substrate [32]. Implanted atoms interact with the point defects, in other words, oversized atoms with vacancies and undersized atoms with interstitials. Hence, they are moved out to the nearby point defect sinks, such as the surface or interfaces, when the substrate temperature increases sufficiently to move the related point defects, that is, vacancies for oversized atoms and interstitials for undersized atoms. For example, in the case of oversized atoms, they are hardly made to implant when the substrate temperature is high enough to move the vacancies. From Fig. 13c it is seen that a number of voids

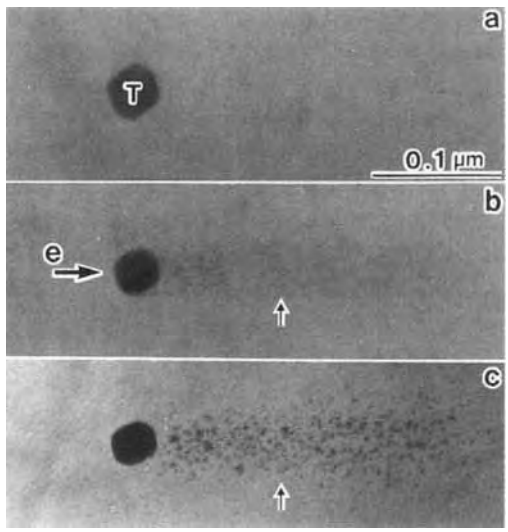


Figure 13. HVEM images taken (a) before and (b) after implantation of Pb atoms into an Al Crystal at 2 MV and 175 K. (c) Image taken after post-annealing at 573 K for 36 ks [40].

appear together with ultrafine Pb particles in the implanted regions after postannealing [32]. This means that aluminum vacancies induced by the electron irradiation tightly interact with implanted Pb atoms at low temperatures at which the vacancies do not move. This is the reason why the lattice distortion of the Al matrix, resulting from a large atomic size difference, is markedly relieved by the buffer effect of the vacancies, as recognized from slight increase of the diffraction contrast in implanted regions in Fig. 13b.

Furthermore, it is emphasized in Fig. 13 that the implanted regions are formed exactly along the incident direction of electrons so that the implantation occurs inside a cylinder parallel to the incident beam direction and having a cross-section corresponding to the size of preexisting Pb particle. This is due to electron channeling occurring in the Al matrix during the implantation [32].

To reiterate: HVEMs has great advantages when compared with other methods using ions and neutrons; they are an indispensable tool for making various new nonequilibrium phases by amorphization and/or foreign atom implantation.

1.4.5 Conclusions

The present review has emphasized that electron channeling by simultaneous reflection becomes very important at high voltages, and that various specimen treatment devices have also been improved as well as stabilization of the high accelerating voltage of the electrons. As a result, not only in situ experiments of various phenomena occurring in materials but also the formation of new functional materials as nonequilibrium phases have been carried out precisely by high voltage electron microscopy. Stable high energy electrons, obtained with HVEMs, can also be applied to various research fields such as ultrafine fabrication and the generation of synchrotron orbital radiation as well as to the construction of compact HVEMs for high-resolution structure imaging, precise analytical microscopes, etc.

One example in which HVEMs have shown a great advantage has been in the recent investigations on the anomalous behavior of atom clusters in the nanometer range as a new type of nonequilibrium phase [39].

Finally, because of the importance of information on valence electrons in materials science, there are many advantages to be accrued by combining HVEMs with other techniques such as Auger electron spectroscopy or STS. This will undoubtedly provide a fruitful path for the future.

1.4.6 References

[1] W. Bollman, *Phys. Rev.* **1956,** *103*, 1588.
[2] P. B. Hirsch, R. W. Horn, M. J. Whelan, *Phil. Mag.* **1956,** *1*, 677.
[3] H. Fujita, Y. Kawasaki, E. Furubayashi, S. Kajiwara, S. Taoka, *Jpn. J. Appl. Phys.* **1967,** *6*, 214.
[4] H. Fujita, *J. Phys. Soc. Jpn.* **1966,** *21*, 1605.
[5] H. Fujita, H. Yamada, Proceedings of the International Conference on the *Strength of Metals and Alloys, Jpn Inst. of Metals, Tokyo, Japan* **1968,** p. 943.
[6] H. Fujita, *J. Electr. Microsc. Techn.* **1982,** *3*, 243.
[7] G. Dupouy, F. Perrier, *J. Microscopie* **1962,** *1*, 167.
[8] H. Hashimoto, *J. Appl. Phys.* **1964,** *35*, 277.
[9] K. Kobayashi, M. Ohara, Proceedings of the 6th International Conference on *Electron Microscopy*, Kyoto, Japan **1966,** *1*, 579.

[10] F. Fujimoto, S. Takagi, K. Komaki, H. Koike, Y. Uchida, *Radiation effects* **1972,** *12*, 153.
[11] A. Howie, *Phil. Mag.* **1966,** *14*, 223.
[12] C. J. Humphreys, L. E. Thomas, L. E. Lally, R. M. Fisher, *Phil. Mag.* **1971,** *23*, 87.
[13] F. Nagata, K. Fukuhara, *Jpn. J. Appl. Phys.* **1967,** *6*, 1233.
[14] D. Watanabe, R. Uyeda, M. Kogiso, *Acta Cryst.* **1968,** *A24*, 249.
[15] M. Kiritani, *Ultramicroscopy*, North-Holland **1991,** *39*, 135.
[16] K. Urban, A. Seeger, *Phil. Mag.* **1974,** *30*, 1395.
[17] P. B. Hirsch, *J. Phys. Soc. Jpn.* **1962,** *B-II*, 143.
[18] H. Fujita, N. Sumida, T. Tabata, *J. Phys. Soc. Jpn.* **1973,** *35*, 224.
[19] H. Fujita, *J. Electr. Microsc. Techn.* **1989,** *12*, 201.
[20] H. Fujita, *JSIJ International* **1990,** *30*, 70.
[21] E. P. Butler, *Rep. Prog. Phys.* **1979,** *42*, 833.
[22] *High Voltage Electron Microscopy*, Academic Press, London and New York (Ed: P. R. Swann, C. J. Humphreys, M. J. Goringe) **1973.**
[23] Proceedings of the 4th International Conference on *High Voltage Electron Microscopy*, Toulouse, France **1975.**
[24] Proceedings of the 5th International Conference on *High Voltage Electron Microscopy*, Kyoto, Japan **1977.**
[25] Proceedings of the 6th International Conference on *High Voltage Electron Microscopy*, Antwerp, Belgium **1980.**
[26] Proceedings of the 7th International Conference on *High Voltage Electron Microscopy*, Berkeley, USA **1983.**
[27] Annual Proceedings of the Electron Microscopy Society of America, San Francisco, CA **1983.**
[28] R. Sandstrom, B. Lehtinen, R. Lagneborg, *Scand. J. Metallurgy*, **1975,** *4*, 1741.
[29] E. P. Batler, K. F. Hale (Eds), *Dynamic Experiments in the Electron Microscope, Practical Methods in Electron Microscopy*, North-Holland, Amsterdam **1981.**
[30] Proceedings of the 1st Osaka International Symposium on HVEM, *In Situ Experiments with HVEM*, Research Center for UHVEM, Osaka University, Osaka **1985.**
[31] Proceedings of the 2nd Osaka International Symposium on HVEM, *New Directions and Future Aspects of HVEM; Ultramicroscopy*, North-Holland, Elsevier Science Publishers B.V. **1992.**
[32] H. Fujita, *Materials Trans. JIM* **1990,** *31*, 523.
[33] H. Fujita, C. Lu, *Materials Trans., JIM* **1992,** *33*, 892; 897.
[34] H. Saka, K. Sasaki, T. Ohashi, I. Ohtsuka, T. Kamino, M. Tomita, Reference [31] **1992,** 110.
[35] K. Shimizu, K. Otsuka, A. Nagasawa (Eds), Special Issue on Matensitic Transformations, *Materials., JIM* **1992,** *33*, 165.
[36] T. Imura, H. Saka, N. Yukawa, *J. Phys. Soc. Jpn.* **1969,** *26*, 327.
[37] H. Fujita, N. Sumida, *Hitachi Inst. News* **1986,** *19*, 3.
[38] H. Mori, H. Fujita, Proceedings of the International Symposium on *Non-Equilibrium Solid Phases of Metals and Alloys. Suppl. to Trans. JIM* **1988,** *29*, 93.
[39] H. Fujita, *Materials Trans., JIM* **1994,** *35*, 563.
[40] H. Fujita, H. Mori, Proceedings of the International Symposium on *Non-Equilibrium Solid Phases of Metals and Alloys. Suppl. to Trans. JIM* **1988,** *29*, 37.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

1.5 Convergent Beam Electron Diffraction

1.5.1 Introduction

Convergent beam electron diffraction (CBED) is perhaps best known as a method for determining crystal point and space groups, and, when combined with elemental analysis using X-rays or electron energy loss spectroscopy, it can be used to determine the composition and structure of small crystals. However, CBED has proved a versatile technique with a much wider range of applications. In this chapter we first explain what information is contained in CBED patterns. We then survey some recent developments including the quantitative determination of crystal bonding and atomic coordinates, the phasing of reflections using coherent diffraction and the analysis of crystal defects and interfaces. For more in-depth accounts of the techniques and applications, the reader is referred to three standard texts [1–3].

CBED is now widely accessible, owing to the commercial development of analytical electron microscopes which have small probe optics and vacuum engineering to minimize probe contamination. For a summary of the instrumental requirements for CBED and an explanation of experimental techniques, the reader is referred to the article by Vincent [3]. In the following we refer to two techniques. In conventional CBED, the incident beam is brought to a focus on the specimen, the probe size being chosen such as to eliminate any variation in sample thickness, orientation or structure across the probe diameter. In large angle convergent beam electron diffraction (LACBED), see Fig. 1, the probe focus is appreciably above or below the specimen such that, in the object plane, the straight-through and diffracted beams are brought to foci which are separated laterally. By inserting a small selected area aperture (usually in the range 1–5 μm) a single diffracted beam can be selected such that the resulting diffraction pattern has only one disc. Although this has the apparent disadvantage of almost certainly introducing specimen variations across the incident probe, LACBED has some important advantages over CBED. First, whereas the total convergence angle in CBED is usually limited to avoid disc overlap, to a maximum of about 2×10^{-2} rad, the convergence angle in LACBED may be five to ten times greater, limited only by electron optics. Second, the selected area aperture acts, like an objective aperture in imaging, to filter out inelastic background. Jordan et al. [4] have shown that, by using a

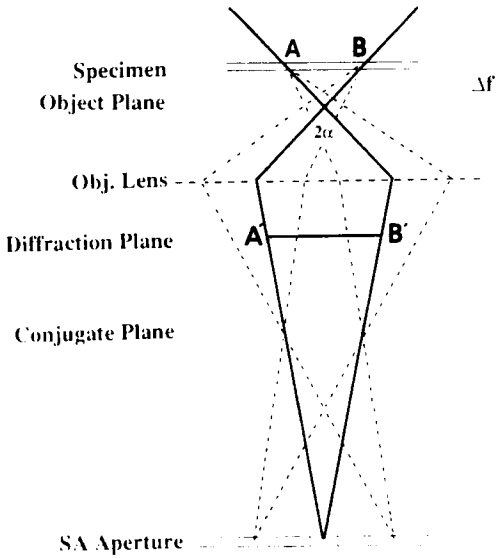


Figure 1. Schematic diagram of LACBED optics with the specimen Δf above the focus of the incident probe.

large defocus of the incident probe (Δf in Fig. 1) and a small selected area aperture (2 μm), the beam acceptance angle may be reduced to less than 10^{-3} rad. Almost all the thermal diffuse background and some of the plasmon scattering can thereby be removed from the resulting LACBED pattern. Finally, in LACBED, since any illuminated region of the specimen sees only a subset of incident beam directions, a shadow image of the illuminated region appears on the final LACBED disk. The spatial resolution of this image is given by the minimum probe size in the limit of geometrical optics. The resulting mixture of real space and diffraction information has proved particularly important in understanding diffraction from defects such as dislocations where rapidly varying strain fields are involved (see Sec. 1.5.7 of this Chapter).

1.5.2 Geometry and Crystal Symmetry Determination

The Ewald sphere construction in the reciprocal lattice is by far the most convenient way of understanding the geometry of electron diffraction [5]. As a result of the very short wavelength of high energy electrons, the radius of the Ewald sphere is very much greater than the spacing between the reciprocal lattice points, typically 50 or 100 times greater. When the electron beam is incident along a zone axis direction of a crystal structure, a very flat Ewald sphere passes very close to a large patch of reciprocal lattice points surrounding the origin of a reciprocal lattice plane (zero layer plane) perpendicular to direction of incidence. With increasing scattering angle, the Ewald sphere eventually curves away from this zero layer plane and intersects successive layers of the reciprocal lattice, parallel to the zero layer plane, in a series of concentric circles of increasing radius G. Thus, the diffraction pattern, viewed at a sufficiently low magnification (camera length) consists of a central patch of reflections with additional large angle diffraction arranged in concentric circles of increasing radius (Fig. 2). The successive layers of the reciprocal lattice, where these reflections appear, are known as Laue zones; the zero layer is known as the zero order Laue zone (ZOLZ) and the others are called higher order Laue zones (HOLZ). The number of HOLZ rings visible depends on the angular view available, the microscope operating voltage and the spacing H of the successive layers of the reciprocal lattice along the zone axis direction. For electrons with wavenumber k we have, to a

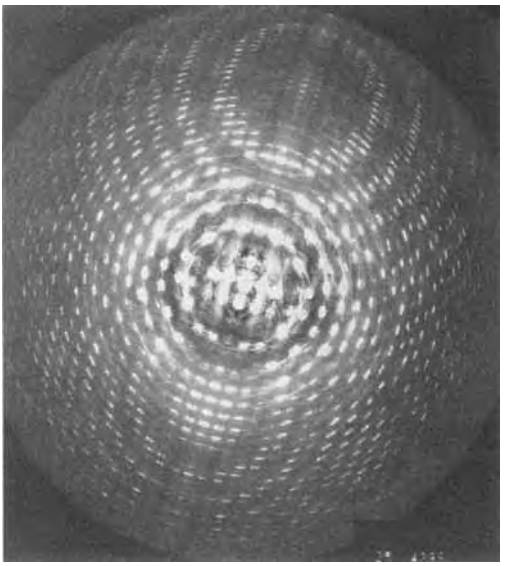


Figure 2. Convergent beam electron diffraction pattern of $NbSe_3$ taken at 120 kV with the specimen cooled by liquid helium. The concentric circles of increasing radius at the center of the pattern are based on the chosen zone axis of incidence. Two other secondary zone axes are clearly visible along the mirror line through the center of the pattern and other zone axes off the mirror line may also be discerned.

good approximation

$$G \approx \sqrt{2kH} \tag{1}$$

In practice, large angle scattering is strongly affected by the Debye–Waller factor and so the visibility of the successive HOLZ rings may be limited by thermal or static disorder in the sample under investigation.

The existence of HOLZ diffraction provides firm evidence of the stacking sequence along the direction of electron propagation. The lack of HOLZ diffraction is an indication of disorder along this direction [6]. This might be planar-disorder as in the case of layer-structured materials with weak interlayer bonding, line-disorder as from a high dislocation density or point-disorder. The relative weakness of HOLZ diffraction is the justification for ignoring it, as in the case of the so-called projection approximation. Neglect of HOLZ diffraction is equivalent to projection of the atomic structure along the zone axis direction. That this is only a first approximation and sometimes a gross oversimplification is evident from Fig. 2. One further point of considerable importance is that the Bravais lattice may be obtained by examination of CBED patterns showing HOLZ diffraction without overlap of diffraction orders [7]. To simplify the analysis it is preferable to perform the analysis at a high symmetry zone axis if possible. A mesh of lines is created in the ZOLZ with nodes at the reflections and a further mesh is constructed from the first order Laue zone (FOLZ) reflections. By the coincidence or not of the nodes of the two meshes the Bravais lattice may be determined.

The nature of HOLZ diffraction is generally rather different from that of ZOLZ diffraction. To a reasonable first approximation it can be treated kinematically, a point that will be discussed in more detail in Sec. 1.5.5 of this Chapter. In this approximation, the width of the diffraction peak, corresponding to a particular reflection $\boldsymbol{g}$, will depend on the magnitude of its associated structure factor F_g. For large values of $\boldsymbol{g}$ as in HOLZ diffraction, F_g is small and so narrow diffraction maxima will occur. The locus of all points within a given cone of electron convergence that satisfy the Bragg condition for that particular reflection is essentially a straight line (in fact it is given by the intersection of two spheres of very large radius, k, one on the origin and one on the reciprocal lattice point $\boldsymbol{g}$ from the origin).

Thus HOLZ diffraction takes electrons away from the cone of incidence along a narrow line perpendicular to $\boldsymbol{g}$ and their Bragg reflection produces a bright line of the same length, width and orientation in the appropriate HOLZ ring. The dark line in the central disk of the CBED pattern is known as a HOLZ deficiency line and the associated bright line in the HOLZ ring is called a HOLZ excess line [8]. Because of their narrow angular width, the HOLZ lines have a number of useful applications. In the case of very fine lines, resulting from a distant FOLZ ring, computations of HOLZ line patterns based on kinematic diffraction theory are likely to be valid with two important reservations [9]. First, for reasons to be discussed in this Chapter, Sec. 1.5.5, the microscope operating voltage must be allowed to vary within about $\pm 3\,\%$ of the apparent value (we shall use the term pseudokinematic to describe HOLZ diffraction on account of this factor). Second, where two lines $\boldsymbol{g}$ and $\boldsymbol{g}'$ intersect at a small angle and $F_{\boldsymbol{g}-\boldsymbol{g}'}$ is large, then strong dynamical coupling of the HOLZ lines corresponding to $\boldsymbol{g}$ and $\boldsymbol{g}'$ will occur near their intersection.

The first use of HOLZ lines is in lattice parameter measurements [10]. One can avoid zone axis directions to reduce the dynamical effects that introduce the pseudokinematic approximation rather than the true kinematic approximation, or else concentrate on changes of lattice parameter caused by temperature or composition changes so that dynamical corrections are unlikely to be important. The second use of HOLZ lines is in strain measurement [11]. Some words of caution are necessary in this context. One of the chief attractions of the technique is the ease with which HOLZ-line distortion may be detected. However, accurate measurements of strains from such observations are rather hard to achieve for a variety of reasons. First, the measurements are being made on thin films where considerable stress relaxation will have occurred and has to be taken into account. Second, the process of making the specimen may have introduced additional stresses that are not immediately obvious (such as are caused by amorphous surface films or surface oxides.) Third, differential expansion effects caused by albeit small beam heating of an interface region between dissimilar materials may lead to spurious results. Finally, to obtain fine HOLZ lines it is generally necessary to have specimen thicknesses of <100 nm and the strain to be measured should remain constant along the column explored by the electron beam, to within the accuracy of the measurements. It is common to achieve measurements of strain or lattice parameters to about 10^{-3} by the HOLZ line method. Greater accuracy down to about 2×10^{-4} can be achieved in favorable cases but energy loss to plasmon excitation then becomes a fundamental limitation.

CBED patterns evidently contain a wealth of symmetry information and this aspect of the subject is now well-understood and widely exploited [12]. The dynamical nature of electron diffraction far from being a complication, is the very strength on which the technique is based. As a result, the absence of inversion symmetry is detectable *straightforwardly*, the symmetry of the patterns may be related directly to the 32 point groups rather than the 11 Laue groups (created when Friedel's law is obeyed, as in normal X-ray diffraction). The theoretical analysis of symmetry

is based on very general and widely applicable approximations of high energy electron diffraction. Tables have been produced that allow one to go directly from experimental observations to point group and space group determinations. In favorable cases this is an extremely efficient process and the information gathered at one zone axis can be sufficient to perform an unambiguous space group determination. More commonly, information has to be gathered at two or more zone axes to complete the analysis. Fortunately, the inclination of the specimen in tilting to a zone axis has, generally, very little influence on pattern symmetry.

The nature of electron diffraction symmetry is more subtle than may be apparent at first. The symmetry of the direct beam may be higher than that of the whole diffraction pattern; individual orders of reflection can have internal symmetry that is important in determining which of the 31 different diffraction groups a particular pattern belongs to. The steps required in identifying the diffraction group and from that the crystal point group are discussed in several previous works [13].

Having determined the crystal point group, the presence of screw axes or glide planes may be inferred by the study of X-ray forbidden reflections in well-oriented zone axis patterns. These appear in electron diffraction as a result of multiple scattering but nevertheless lines of absent intensity occur within these reflections because of exact cancellation between the various contributing diffraction paths. The individual diffraction paths may be paired together such that each path contributes an equal amplitude to the reflection precisely out of phase with the contribution from the other path. The resulting lines of absence are variously called dynamic absences, Gjønnes-Moodie lines, or dark bars. Accounts of their use in space group determination are given in several references [14].

1.5.3 Bloch Wave Theory

To understand the details of CBED patterns it is generally necessary to make reference to a form of dynamical diffraction theory. The simplest of these is the so called two-beam theory where only the incident beam and one diffracted beam are considered significant [15]. According to this theory, ignoring absorption, the intensity of the direct beam, I_0, is given by

$$I_0 = 1 - 4C_0^{(1)^2} C_0^{(2)^2} \sin^2 \frac{\pi t}{\xi_g} \sqrt{1 + w_g^2}$$

$$= 1 - I_g \qquad (2)$$

where I_g is the diffracted beam intensity, $w_g = s_g \xi_g$ is called the deviation parameter from the g Bragg condition, $\xi_g \propto 1/F_g$ is the so-called extinction length, $s_g \sim g\Delta\theta$ where $\Delta\theta$ is the angular deviation from the g Bragg condition and t is the thickness of the specimen.

We can also write

$$I_g = 4C_0^{(1)^2} C_0^{(2)^2} \sin^2 \frac{\pi t}{\xi_g^{eff}} = \frac{\sin^2(\pi t/\xi_g^{eff})}{1 + w_g^2} \qquad (3)$$

where $\xi_g^{eff} = \xi_g / \sqrt{1 + w_g^2}$ and ξ_g^{eff} is called the effective extinction distance. Its value decreases symmetrically for $\pm\Delta\theta$.

The basis of this theory is the Bloch wave formulation of dynamical diffraction.

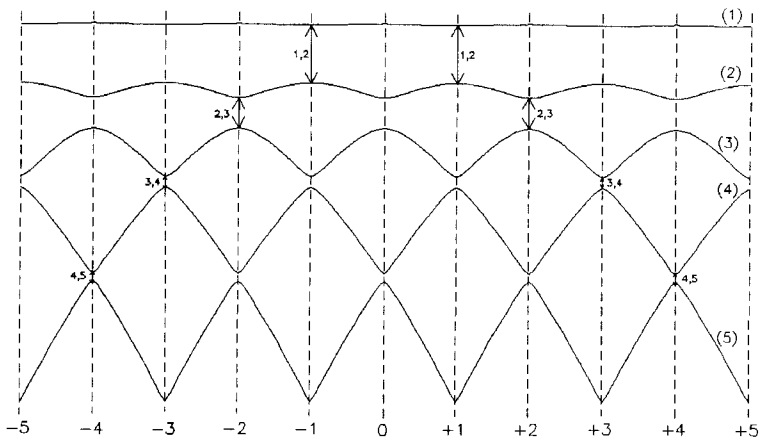


Figure 3. Dispersion surface for a (*n n n*) systematic row of silicon at 300 kV. The dispersion spheres appear as parabolas in this diagram and are asymptotic to the dispersion surface away from Brillouin zone boundaries. These parabolas are centered on each of the reflections *n*. At the *n* Bragg position, the effective extinction distance is the reciprocal of the gap labelled as *n*, *n* + 1. This diagram should be related directly to the LACBED result in Fig. 4.

It is particularly convenient for understanding many of the diffraction effects that occur. The results of this form of diffraction theory are conveniently represented in terms of a so-called dispersion surface constructed from dispersion spheres (of radius proportional to the free electron energy) centered on each of the reciprocal lattice points. The spheres intersect at Brillouin zone boundaries, where electron diffraction occurs, causing standing (Bloch) wave formation and splitting of the constant energy surfaces near the circles of intersection of the spheres (Fig. 3) These effects are the direct equivalent of Fermi surface and forbidden band formation for conduction electrons. $C_0^{(1)}$ and $C_0^{(2)}$ in the expressions above are the Bloch wave amplitudes of the two branches approaching an intersection point of the constant energy spheres.

A particular attraction of the Bloch wave formulation is that only a small number of excitation amplitudes are significant at any given orientation and the same or an even smaller number show significant orientation variation within the angular range of a CBED disk. In the simplest case, which often applies, only two branches need to be considered.

Let us consider the application of two beam theory to the LACBED pattern formed for silicon (1 1 1) systematic diffraction from a wedge shaped crystal (Fig. 4). The thickness of the specimen increases vertically up the page and the position midway between the two sets of curved fringes is the symmetric position with the direction of incidence perpendicular to the (1 1 1) systematic row. The series of dark vertical lines indicate the Bragg positions of higher order reflections (*n n n*). They are marked by integers on the corresponding dispersion surface shown in Fig. 3. At the ±1 Bragg positions the fringes have a maximum spacing that corresponds to the reciprocal of the gap opened up at the first Brillouin zone boundary between branches (1) and (2) of the dispersion surface. At the ±2 Bragg positions, a series of widely-spaced arrow-like fringes are visible. These correspond to the results of another two branch theory, this time for the smaller gap opened up at the second Brillouin zone boundary between branches (2) and (3) of

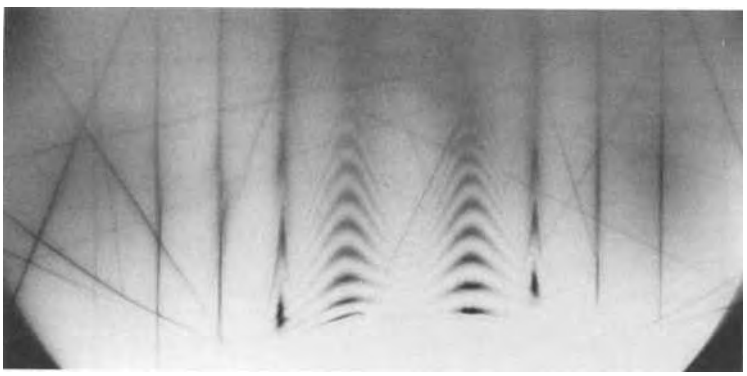


Figure 4. Experimental LACBED wedge/rocking curve for the (*nnn*) systematic row of silicon at 300 kV. Lines inclined to the vertical are nonsystematic reflections, not taken into account in the calculated results shown in Fig. 3. They clearly show the expected lack of vertical mirror symmetry. Thickness fringes may be observed clearly at the ±1, and ±2 Bragg conditions: they can also be detected (just) at the ±3 Bragg condition.

the dispersion surface. Because the gap is smaller, the extinction length is longer and the angular range over which the local two-branch theory applies is narrower than the first case. In the case of the ±3, ±4 and ±5 Bragg conditions the gaps opened up between the relevant dispersion surface branches are so small that very long extinction lengths occur. Hence the first maximum is not reached before absorption takes over, resulting in diffuse scattering at the top edge of the pattern, and only dark lines corresponding to the first extinction minima are observed. The lines become narrower and fainter as n increases.

In addition to demonstrating the value of a series of local two-branch approximations over local orientation ranges, there are several further interesting points about this micrograph. According to kinematical diffraction theory, for spherical atoms with harmonic lattice vibrations, there would be no extinction at the ±2 Bragg positions. The extinction observed is essentially the result of dynamical diffraction along the systematic row, with a small additional contribution from bonding effects and anharmonicity. According to many beam dynamical theory, the extinction length at the ±1 Bragg positions is 103.7 nm and at ±2 it is 190.0 nm. In fact the ratio between the two extinctions obtained from measurements on the micrograph, is closer to 2.5. This difference is believed to be the result of nonsystematic interactions that perturb the purely systematic diffraction assumed in the calculations. The presence of non systematic reflections may be observed clearly in the micrograph and they break the mirror symmetry along the center line of the pattern as expected. The −6 reflection is not observed and it is expected to be very weak because, like the (2 2 2) reflection, it can only appear through systematic dynamical interactions if bonding effects and anharmonicity are ignored.

This same theory is readily adapted to making accurate specimen thickness measurements from fringe patterns in CBED from systematic rows of reflections [16]. The one dimensional example given above is easily adapted to describe two dimensional (zone axis or cross-grating) diffraction [17]. Those Bloch states with high probability density of electrons on the atom strings have low potential energy

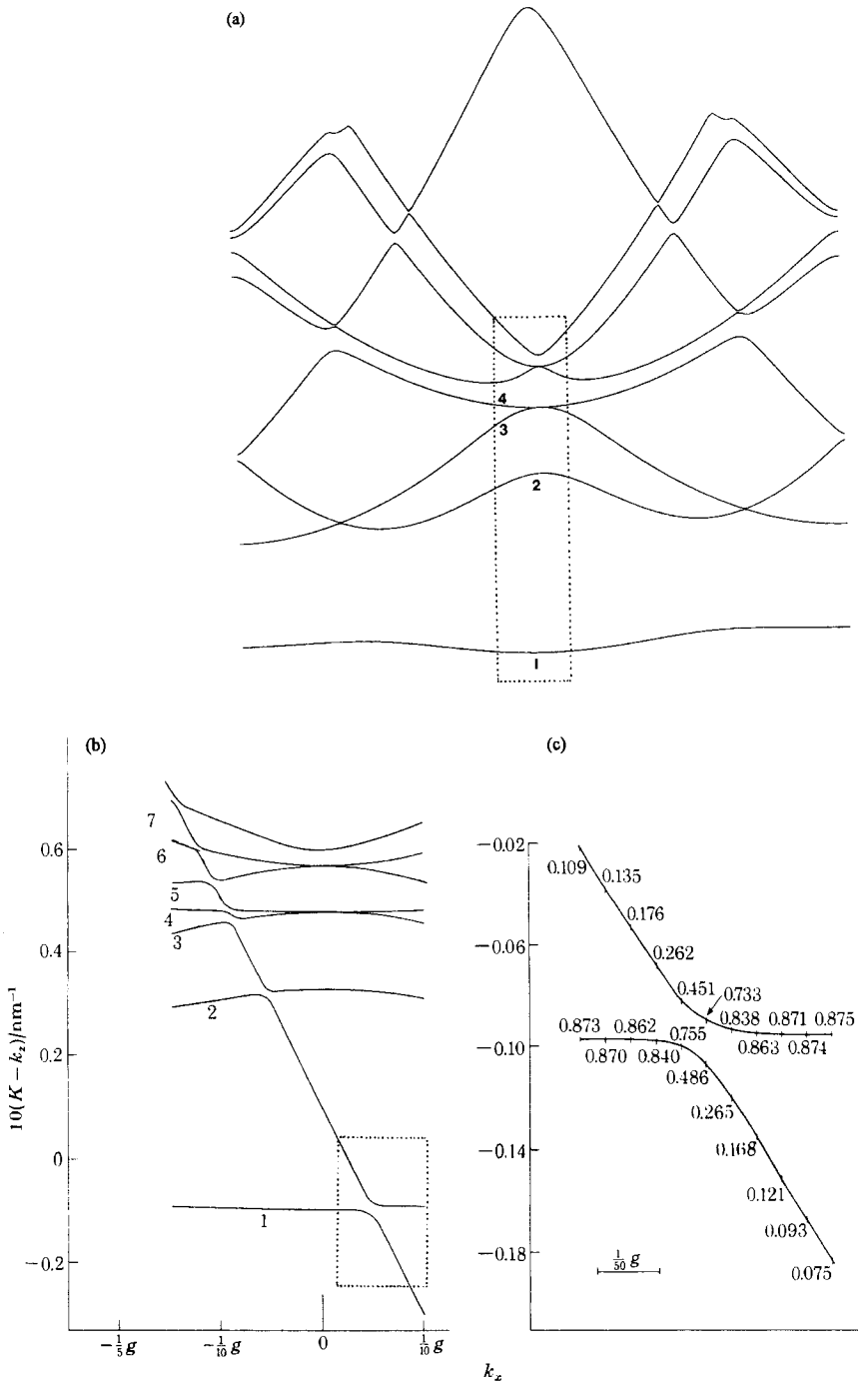


Figure 5. $\{1\,\bar{1}\,1\}$ section through a Si$\langle 1\,1\,1\rangle$ zone axis dispersion surface at 100 kV. The diagram is inverted relative to Fig. 3 in order that the lowest potential energy state (1 s) is at the bottom. (a) Large region of the dispersion surface centered on the origin for the case of two dimensional (zero layer) diffraction, (b) boxed region of (a) is shown in greater detail with a single HOLZ $(\bar{5}\,\bar{5}\,1\,1)$ reflection added to the calculation $(g = 1/a[\bar{1}\,\bar{1}\,2])$. The intersections of the HOLZ sphere with the zero layer branches causes local splitting and hybridization as detailed in (c) for the boxed region in (b).

and high kinetic energy. They represent electron states bound in the atom-string potential wells which, in the simplest case, have cylindrical symmetry. These bound-states are analogous to atomic bound-states and have principal and angular momentum quantum numbers nl where $l \leqslant n-1$ (e.g., 1 s, 2 p). The states are ordered in the atom-string potential first by their principal quantum number, then by their angular momentum quantum number, the higher the l value the lower the energy of the state. It follows that the two lowest-energy bound-states are 1 s and 2 p. However, except in the case of rather deep atom string potential wells, there is only one bound state (1 s) and the nearly free states, just above the top of the well, have a different order, with 2 s below 2 p in energy (Fig. 5). This latter situation is very common and gives rise to the concentric ring pattern often observed at zone axes of simple materials. The rings are produced by two-branch interference between the 1 s and 2 s states exactly analogous to the state of affairs described by Eq. 2. The wells may be made deeper by a number of factors and it is possible [18] to define a string-strength parameter S that includes these factors to a first approximation

$$S = \frac{\gamma Z A}{d} \tag{4}$$

where γ is the relativistic mass factor, Z is the mean atomic number in the repeat distance, d, along the atom string and A is the area of the Wigner–Seitz cell associated with the atomic string. The γ factor arises because we can regard the relativistic effect, on increasing the microscope operating voltage, as multiplying the potential that the electron experiences by the increase of its relativistic mass.

$$\left(\frac{-h^2}{2m_0^2}\nabla^2 + \frac{m}{m_0}V(\boldsymbol{r})\right)\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r}) \tag{5}$$

It follows from this argument that an atom string potential well that is not quite deep enough to hold two bound states at 100 kV may be able to do so at some higher operating voltage, in consequence of the increase of relativistic mass [19]. For this to happen interchange of the order of the 2 s and 2 p states has to occur and it does so through a so-called accidental degeneracy (i.e., not symmetry determined) just at the point where the 2 s level touches the top of the string potential well. This accidental degeneracy is known as the critical voltage effect and the voltage at which the degeneracy occurs may be determined experimentally with great accuracy [20]. At higher voltages the two $n = 2$ states are bound and reversed in order, 2 p below 2 s. There is an exactly equivalent critical voltage effect in one dimensional electron diffraction but these critical voltages tend to occur at rather high microscope operating voltages [21].

Three-dimensional diffraction involves a relaxation of the projection (atomic string) approximation and requires the addition of free electron dispersion spheres centered on HOLZ reflections. It is frequently valid to consider the effect of a particular HOLZ reflection g in isolation from other HOLZ reflections by adding just one sphere centered on the particular HOLZ reciprocal lattice point to the dispersion surface construction [22]. Because of the large distance of g from the zone axis the g dispersion sphere intersects the zero layer dispersion surface at an appreciable angle and wherever it cuts a particular zero layer branch splitting occurs [9] and

a new two-branch situation is created (Fig. 5). It is precisely because the HOLZ line that results from the intersection between the HOLZ dispersion sphere and the branch of the zero layer dispersion surface is displaced from the kinematical position that the term pseudokinematical was adopted earlier and a degree of voltage variation had to be introduced in computer matching of HOLZ-line patterns. The splitting introduced between the zero layer branch and the HOLZ dispersion sphere is rather small [9] corresponding to an extinction length (ξ_H) of several hundred nanometers. It follows that we can often make the kinematical approximation to Eq. (2), that is,

$$\sin\frac{\pi t}{\xi_H} \approx \frac{\pi t}{\xi_H}$$

In this sense, therefore, the HOLZ diffraction is kinematic and also in the sense that the splitting that occurs may be calculated by perturbation theory [22]. However, the displacement of the zero layer branch from the dispersion sphere on the origin is a dynamical effect that depends on the strength of the atom-string potential. In this sense HOLZ diffraction is not kinematical. Moreover, there may be several appreciably excited branches of the zero-layer dispersion surface, each giving a separate HOLZ line, displaced from, but approximately parallel to, the others. This HOLZ line splitting or fine structure is also a purely dynamical effect, permitting direct visualization of the strongly excited zero layer branches. Moreover, the extinction length ξ_H may be greatly reduced (by something like a factor of five) from the value deduced from kinematical theory ($\xi_H^{kin} \propto 1/F_H$), where F_H is the HOLZ reflection structure factor) and hence there is another sense in which HOLZ diffraction is not kinematical. Indeed, if it were not for the zone axis reduction of the HOLZ extinction lengths, HOLZ lines would probably not be visible in zone axis CBED patterns.

1.5.4 Crystal Structure Determination by Convergent Beam Electron Diffraction

For thin films of crystals which contain principally light elements, structure determination by plane wave electron diffraction has been based upon the techniques developed for the analysis of kinematic X-ray data [23]. Recent advances include the introduction of direct methods and maximum entropy criteria to phase high order reflections beyond the point resolution limit in lattice images [24]. However, this approach is not reliable for crystals which contain heavier atoms, where many of the low order reflections are associated with extinction distances comparable to the crystal thickness, even for CBED patterns acquired with the beam focused onto a crystal with a local thickness of a few nanometers. A partial solution, at least for inorganic crystals of moderate complexity, is offered by analysis of HOLZ reflections, which intercept the Bragg condition in zone axis CBED patterns, usually have extinction lengths much larger than the crystal thickness, are not located within strong systematic rows, and are associated with structure factors sensitive to small changes in the atomic parameters.

The basis for a quantitative analysis of HOLZ intensities is given by the Bloch

wave formalism for the solution of the Schrödinger equation in the projection approximation. Along major zone axes, the exit wave function is a phased sum over relatively few excited Bloch states, often described as molecular combinations of localized two-dimensional atomic states. The eigenvalues associated with a near-degenerate molecular cluster of 1 s type Bloch states localized on a symmetry-related set of atom strings in the projected potential may be quite distinct from the eigenvalues of other localized or channeling Bloch states. It follows that the components of the electron wave vector along the zone axis also differ, and that the inner excess lines observed within the disks of HOLZ reflections correspond to diffraction from Bloch state clusters localized on subsets of atom strings, with a related shift of excess and deficiency lines away from their kinematic positions.

The Bloch wave formalism is equally useful for constructing a model of the factors which control the relative intensities of HOLZ excess lines. In the kinematic approximation, the amplitude of a HOLZ reflection is proportional to the relevant structure factor of a two-dimensional conditional potential, $U^n(\boldsymbol{R})$, constructed by a Fourier sum over structure factors in layer n of the reciprocal lattice, where $\boldsymbol{R}$ is a position vector within the projected cell. For $n = 0$, $U^0(\boldsymbol{R})$ is the usual projected potential, which is real by definition, but the general conditional potentials are complex with phase factors proportional to atom positions along the projection axis, being simply an expression in real space of the information implicit in the structure factors. The final step is to consider the effect on the HOLZ amplitudes of diffraction not from the incident plane wave, but from a cluster of Bloch states. It may be shown [25] that a modified version of the conditional potential is retained, defined by multiplication of $U^n(\boldsymbol{R})$ with the relevant Bloch state at every position within the projected cell. In a quasikinematic limit, the amplitudes of associated HOLZ excess lines remain proportional to the Fourier coefficients of the modified potential which may include contributions from only a single subset of atoms.

Provided that an approximate structure is available for calculation of the zone axis Bloch states, the intensities of suitable HOLZ excess lines may be used to refine the atomic parameters. To improve the statistical accuracy, it is important to combine data acquired over a wide voltage range, equivalent to an annulus of HOLZ reflections. The atomic parameters are refined by standard X-ray procedures, which minimize the mean squared difference between observed and calculated intensities based upon a kinematic expression derived from the partial structure factors. Typical reliability factors for HOLZ refinements (defined as the mean fractional difference between observed and calculated intensities) are larger than equivalent X-ray values although the standard deviation of the atomic parameters is impressively small, often equivalent to an accuracy of 0.001 nm. The explanation for this apparent dichotomy is that only the high order reflections are refined. The intensities of individual reflections are very sensitive to small changes in the atomic parameters but are also subject to perturbation by coupling between adjacent HOLZ reflections that is not included in the quasikinematic approximation described above.

Examples of structure refinement [26–28] include the analysis of displacive

superlattices where the symmetry of the subcell is broken to produce a domain structure not easily accessible to X-ray or neutron techniques. A further application is to metastable crystal structures produced either by interface reactions or by rapid quenching of liquid or vapor phases. For crystals which have no resemblance to known phases, there remains the basic problem of deducing the structure. Some progress has been made by adapting the arguments outlined above for HOLZ diffraction. If the inner excess lines show strong variations of their relative intensities around a major axis, then it is reasonable to assume that the relevant Bloch states are localized, and that the diffracted intensities may be analyzed by X-ray procedures.

Patterson transforms of the HOLZ intensity data were used to deduce interatomic vectors and subsequently refine the atomic parameters in metastable alloys [29, 30]. The correlation peaks were very sharp, being the transform of a restricted set of reciprocal lattice vectors with short wavelengths. This approach fails when the peaks are not visible above the background ripple generated by the truncated data set, combined with the inevitable noise associated with dynamical coupling between HOLZ reflections. A further problem common to all analyses based on localized Bloch states is that weak atom strings, often corresponding to projections of atoms with low atomic number, are not detected. An alternative approach based upon an electron-optical equivalent to the X-ray precession geometry has been explored where the Ewald sphere is rocked through the Bragg condition for an annulus of HOLZ reflections [31]. An example for a rare earth pyrogermanate is shown in Fig. 6, where the Er and Ge positions agreed with the X-ray data, but the oxygen positions were obscured by nonsystematic perturbations.

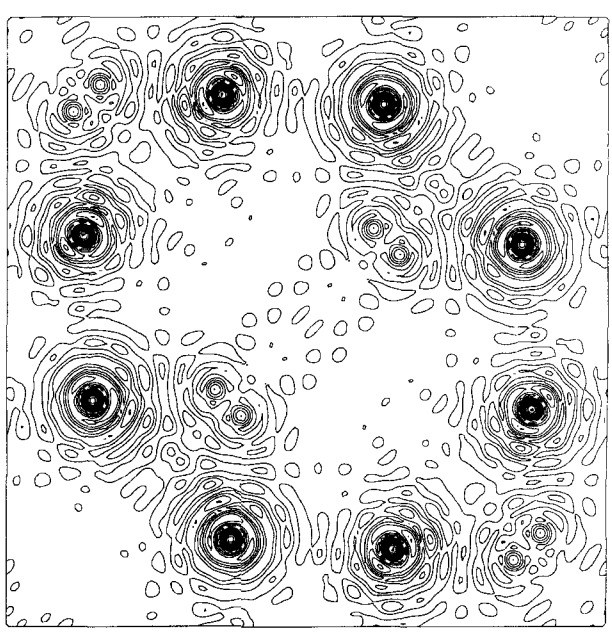


Figure 6 . Example of the projected Fourier transform of a precession pattern for $Er_2Ge_2O_7$ (space group $P4_{1,3}2_12$) calculated from the intensities of an annulus of FOLZ reflections excited around the *c* axis. The peaks corresponding to Er and Ge atoms in general (8-fold) positions were refined from the experimental data. The background ripple around and between peaks is associated with truncation of the Fourier series when summed over the FOLZ reflections. (Courtesy of M. E. Sleight.)

1.5.5 Quantitative Convergent Beam Electron Diffraction

Unlike the analysis of kinematic X-ray and neutron diffraction data, there exists no general method for the inverse solution of the Schrödinger equation to retrieve a complete set of Fourier coefficients, V_g of the crystal potential. Nevertheless, there are several strong incentives, both theoretical and practical, for the quantitative comparison of calculated and experimental CBED intensities. These include the introduction of field emission sources with high brightness and also commercial electron spectrometers to remove the inelastic component from diffraction patterns. When combined with the absence of any requirement for perfect crystals larger than nanometer dimensions, and the continued decrease in the cost of dedicated computers, it has become possible to match many-beam dynamical calculations with experimental CBED patterns, treating the structure factors, thermal parameters and thickness as free variables which are adjusted to minimize the disparity between experiment and theory. In principle, we may infer the composition, atom positions, vibrational amplitudes and the bond charge distribution from a complete knowledge of the crystal potential or the equivalent charge distribution calculated via Poisson's equation.

However, the time required for each dynamical refinement cycle increases rapidly when more beams are included, and has imposed an effective division between structure determination by traditional methods applied to quasikinematic estimates of intensities for a large number of high order reflections (see Sec. 1.5.4 of this Chapter), and the exact calculations applied to simple inorganic crystals with known structures. Even for these crystals, the majority of structure factors are reliably estimated by calculation from the form factors for neutral isolated atoms, and only the phase and amplitude of the lowest order components in the crystal potential are varied to seek a global minimum in phase space between observed and calculated intensities, subject to a statistical χ^2 criterion for acceptable agreement between experiment and theory.

As discussed in detail by Spence and Zuo [2], the purpose of this comparison is to measure with high accuracy the lowest order coefficients of the crystal potential, which are slightly perturbed by the redistribution of charge associated with bonding between atoms. A further advantage of using electrons for measurement of bond charge distributions is that the necessary conversion from voltage to charge reduces the statistical error, which is a vital consideration because bonding represents only a small fraction of the natural overlap between the valence wave functions when atoms are assembled into a crystal. The corresponding maximum change in the lowest order electron structure factors is less than 1 %. Problems which are not yet fully addressed include the treatment of thermal diffuse scattering which remains unfiltered from experimental patterns, the accurate measurement of thermal Debye–Waller factors, which appear in the conversion from voltage to charge distributions, the improvement of models for the imaginary component of the crystal potential which simulates the loss of electrons by inelastic processes, and the accurate estimation of the local composition in compounds.

Measurement of the complex structure factor for a single reflection is based upon the two-beam diffraction geometry, where all other higher order reflections with significant excitation are included in calculations. Alternatively, dynamical interactions between low order reflections perturbed by bonding are enhanced within a single zone axis pattern, and may represent a more efficient strategy for refinement of the structure factors [32] (Fig. 7). As yet, published results have been confined to a few semiconductors, alloys and oxides, but the best quality data agree closely with the extensive X-ray measurements on Si and Ge, and are consistent with band structure calculations. A separate parameter not measured by pattern matching is V_0, the mean crystal potential which is a sensitive measure of charge redistribution in crystals. For self-consistency, measurements of structure factors and thermal parameters must agree with independent estimates derived from thickness-independent contrast features in CBED patterns, notably local (three-beam) coupling between nonsystematic reflections and critical voltage contrast associated with accidental degeneracy of Bloch state eigenvalues.

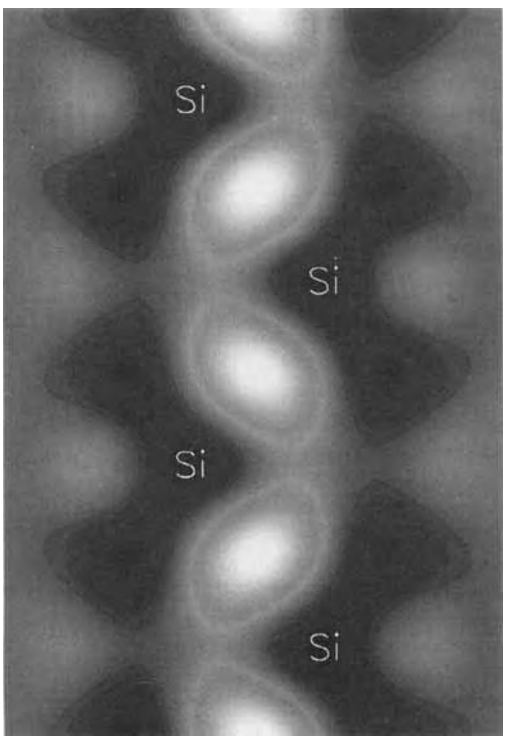


Figure 7. Charge distribution in crystalline Si at 300 K, calculated by matching calculations to a ⟨110⟩ CBED pattern, and seen through a (1 1 0) planar section intercepting the nuclei and covalent bonds. To enhance visibility, the neutral atom charge distribution has been subtracted, leaving bonds as regions of excess negative (light tones) charge, balanced by dark electron-deficient regions, centered on the atoms and also extended into tetrahedral lobes in opposition to the bonds.

1.5.6 Coherent Convergent Beam Electron Diffraction

Provided that adjacent disks do not overlap, CBED patterns from a perfect crystal with plane parallel surfaces normal to the beam are indistinguishable, being independent not only of the probe size, position and defocus, but also of the probe coherence. The latter result follows from the absence of interference between adjacent beams within any disk of a CBED pattern. The situation is entirely different if the convergence angle is increased so that the discs overlap. The random phase relations between beams emitted by the conventional electron source used in most transmission microscopes ensures that only the intensities of adjacent discs are added in the overlap regions, whereas the overlap intensities in the CBED pattern formed by

an ideal coherent source are determined by the phased addition of the diffracted amplitudes. Further, the contrast within overlaps is dependent on the probe position within the projected cell. If the probe is scanned across a set of lattice planes with spacing d and reciprocal lattice vector $g = 1/d$, then the relative phase at any position in the overlap of a pair of reflections related by the difference vector $\boldsymbol{g}$ cycles through 2π in distance d with a corresponding periodic variation of the interference contrast.

For beam convergence angle 2α, adjacent discs overlap if $2\alpha < \lambda/d$, equivalent in real space to the condition that the size of an ideal, diffraction-limited focused probe must be comparable with or less than the planar spacing. This result follows by application of the Airy criterion, where $\Delta = 0.61\lambda/\alpha$, or $\Delta < 1.22d$. When the probe is defocused by Δf relative to the specimen plane, the beam remains coherent but a regular phase variation of the beam is introduced across the lattice planes, and sinusoidal contrast with angular period $d/\Delta f$ is observed in the overlaps, equivalent to interference contrast between virtual diffracted images of the probe.

For a field emission source imaged onto the specimen by electron lenses, the probe size is increased by other factors, including the physical (incoherent) source size, instrumental instabilities and spherical aberration. Until recently, observations of coherent CBED contrast have been limited principally to scanning transmission microscopes equipped with field emission sources. As discussed by Cowley [1] and Spence and Zuo [2], there is a direct relation between the conditions for observation of coherent diffraction contrast discussed above, and the corresponding resolution of lattice fringes in the signal collected by a detector when the probe is scanned across the specimen. Applications to materials science include the detection of site-symmetry elements in nanometer-sized particles and the analysis of fault vectors on a similar length scale. More generally, there are prospects for 'super-resolution' based upon Fourier analysis of the diffraction contrast from a scanned coherent probe in the limit of a thin phase object [33]. In a related development, coherent zone axis patterns (Fig. 8) obtained from conventional transmission microscopes equipped with field emission sources [34, 35] offer the prospect of enhanced resolution in the diffraction plane by stepwise phasing of reflections

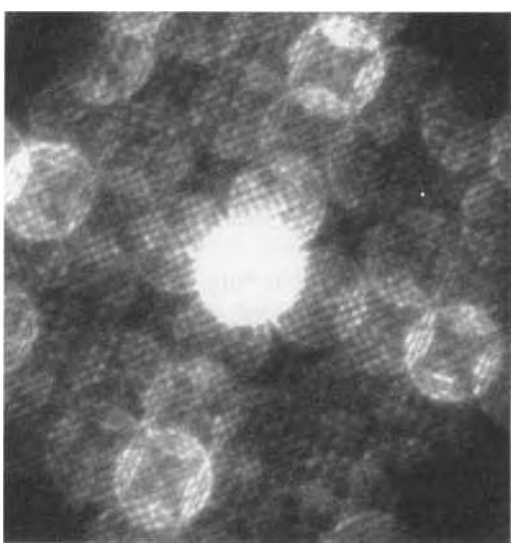


Figure 8. Coherent zone axis CBED pattern for $BaCuO_2$ (space group Im3m, $a \approx 1.3$ nm), aligned with the beam parallel to a cube axis. The convergence angle is adjusted so that CBED discs overlap with first and second nearest neighbors, and the probe is slightly defocused to show interference fringes in both sets of overlaps. (Courtesy of Dr. P. Spellward.)

beyond the limit on point resolution set by spherical aberration. The close relation between coherent diffraction and holography is confirmed by the observation of interference fringes in CBED disks overlapped by an electron biprism [36], where the limits on probe size and coherence are relaxed.

1.5.7 Studies of Imperfect Crystals

In the last few years it has become clear that CBED, and more particularly LACBED, are powerful methods of studying crystal imperfections. The essential point is that imperfections influence the crystal rocking curve. This can be seen most easily by considering the kinematical expression for the two-beam diffraction function,

$$\phi_g(s) \propto \int_0^t F_g(z) \exp[-2\pi i(sz + \boldsymbol{g} \cdot \boldsymbol{R})]dz \tag{6}$$

where F_g is the structure factor, assumed to be depth-dependent, $\boldsymbol{R}$ is any displacement field due to, for example, a dislocation or an interface, s may include any uniform crystal strain and t is the sample thickness along the beam direction. CBED and LACBED allow us to profile the rocking curve $I_g = |\phi_g|^2$ for a selected reflection. Since diffraction angles are small and the Ewald sphere is nearly flat, the rocking curve, at least for low order reflections, is scanned approximately along the incident beam direction. The resulting rocking curve is mapped into the appropriate diffraction disc.

To analyze imperfect crystals it has proved most useful to examine two-beam rocking curves. Extensive studies have been carried out on semiconductor multilayer structures. A review of this topic is found in the literature [37]. Studies of plan view samples have proved most effective since, in this geometry, the essential diffraction information lies along the specimen normal around each fundamental reflection, the direction to which we are most sensitive when samples are flat. Figure 9 shows an example for a periodic multilayer where the rocking curve reveals superlattice satellite reflections, $n\mathbf{q}$, due to variations in $F_g(z)$. The large number of satellite reflections visible underlines both the efficient filtering of inelastic background achieved in LACBED, and the elimination of spatial averaging which would attenuate higher order satellites [37]. The weakness of the satellites at $n \approx 5$, 10, 15 can be related to the ratio of adjacent layer thicknesses [38].

The rocking curves from periodic multilayers can be understood in some detail using kinematical theory since most

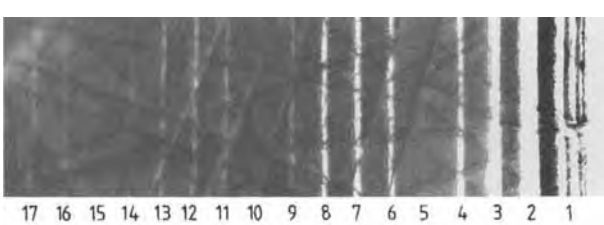


Figure 9. 200 dark field LACBED pattern from a 9.5 nm $In_{0.53}Ga_{0.47}As$/35 nm InP plan view superlattice sample showing superlattice sidebands (numbered). (After R. Vincent et al. [38].)

information is at large deviation parameters. Layer thicknesses and periods can, for example, be determined to near monolayer accuracy. However, for a quantitative fit, dynamical theory is required. Very detailed agreements between dynamical simulations and experiment have been achieved, revealing, in particular, some structure-sensitive systematic row interactions [39, 40].

LACBED has also been used extensively to profile single buried layers in semiconductors using the plan view geometry. In this case, information on the buried layer appears as a modulation of the single crystal oscillations described by Eq. (3). Both layer composition and strain can be profiled, the latter generating asymmetry in the rocking curve between $\pm s_g$. Single atomic steps can be detected [41] and layers down to one or two monolayers analyzed [42]. By combining LACBED with complementary imaging methods, a detailed picture of layer topology at the atomic level can be derived.

LACBED has now been used to profile strains in a variety of structures. In strained multilayers the effect of strain is to cause a rotation of crystal planes inclined to the interfacial plane. The relative rotation between planes in adjacent lattice matched layers is given approximately by

$$\delta\theta = \frac{\eta(1+\nu)}{2(1-\nu)} \sin 2\theta \qquad (7)$$

where η is the natural mismatch, ν is Poisson's ratio and θ is the unstrained angle of inclination of the diffracting planes [43]. In reciprocal space this corresponds to a splitting of common reflections along the growth direction (see Fig. 10). This splitting can be observed in plan-view specimens if the sample is tilted to select reflections from inclined planes. The method has also been applied to cross-sectional samples. Although LACBED is not sensitive to this splitting when the layers are edge-on, since the splitting is then perpendicular to the incident beam direction, a rotation of a few degrees from the edge-on orientation gives a component of the spot separation along the beam direction. This is observed as an angular displacement between diffraction contours. Figure 10 shows an example from recent work on Si/SiGe [44]. Quantitative studies are possible; interpretation can be complicated by surface relaxation in the cross-sectional geometry, although, in

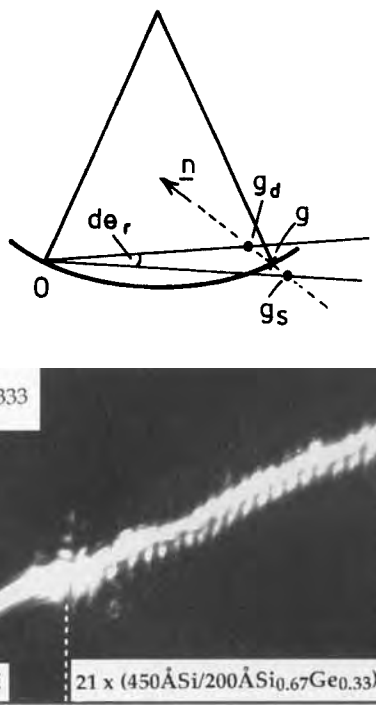


Figure 10. Layer strains in a Si/SiGe superlattice observed by the displacement of a (3 3 3) contour. The schematic diagram shows how the diffraction vectors $\boldsymbol{g}_d$ and $\boldsymbol{g}_s$ from the Si and SiGe layers are displaced along the specimen normal $\boldsymbol{n}$.

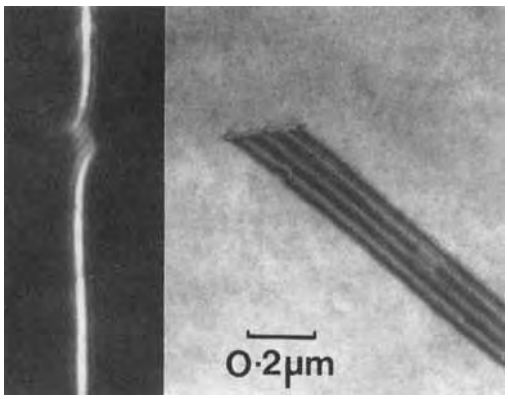


Figure 11. A dark field LACBED pattern taken in the reflection $\boldsymbol{g} = (0\,\bar{1}\,0\,2)$ showing the image of a partial dislocation bounding a stacking fault in CdMnTe. The fault itself has no effect on the contour ($\boldsymbol{g} \cdot \mathbf{R}$ = integer). (After D. Cherns et al. [48].)

principle, such effects can be taken into account [45].

LACBED has proved particularly successful as a method of analyzing dislocations. It was first shown by Cherns and Preston [46] that relatively high order diffraction contours dissociate on crossing dislocations. An example is shown in Fig. 11. Under two-beam conditions, the number n, of subsidiary maxima in a dark field pattern, or subsidiary minima in the corresponding bright field pattern, is simply given by

$$\boldsymbol{g} \cdot \boldsymbol{b} = n + 1 \tag{8}$$

for integer $\boldsymbol{g} \cdot \boldsymbol{b}$ values. Moreover, the direction of the displacement depends on the sign of $\boldsymbol{b}$. Since the characteristics of the dislocation images depend principally on the long-range displacement of the diffracting planes, they are relatively insensitive to imaging parameters such as dislocation depth and foil thickness, and to the presence of impurities near dislocation cores. Breakdowns of the rule do occur where dislocations are close to the foil surfaces and under strongly dynamical conditions [47]. The LACBED method of analyzing dislocations has proved particularly attractive because it gives both the magnitude and the sign of the Burgers vector which are difficult to obtain by imaging methods. Since high order contours can be used, the method is particularly suitable for partial dislocations with small Burgers vectors [48]. A key factor is the choice of reflections which put adjoining faults out of contrast. This is done in fig. 11 which shows part of the analysis of a partial dislocation terminating a stacking fault in CdMnTe. The reflection $g \sim (0\,\bar{1}\,0\,2)$ puts the fault out of contrast ($\boldsymbol{g} \cdot \boldsymbol{R}$ = integer); Eq. (8) can then be applied directly giving $\boldsymbol{g} \cdot \boldsymbol{b} = 4$ (i.e., $n = 3$). The result is consistent with a Burgers vector of $\boldsymbol{b} = \frac{1}{3}(1\,\bar{1}\,1)$.

Although dislocations introduce a strain field which varies with depth z, a stacking fault or interface can introduce a rigid body displacement $\boldsymbol{R}$. It can easily be shown that a fractional $\boldsymbol{g} \cdot \boldsymbol{R}$ in Eq. (6) introduces an asymmetry into the rocking curve, which is reversed when the sign of is reversed. This was used by Morniroli and Steeds [49] as a quick and reliable method of distinguishing intrinsic and extrinsic faults in Si, through the observation of HOLZ deficiency lines in a conventional CBED pattern. The method has been extended to the analysis of interface displacements, notably $CoSi_2$/(1 1 1)Si, where a more quantitative dynamical analysis was applied [50].

Acknowledgements

One of us (JWS) wishes to thank Dr. Martin Saunders for assistance in preparing Figs. 3 and 5.

1.5.8 References

[1] *Electron Diffraction Techniques*, Vol. 1 (Ed.: J. M. Cowley), Oxford University Press, Oxford **1992**.

[2] J. C. H. Spence, J. M. Zuo, *Electron microdiffraction*, Plenum Press, New York, **1992**.

[3] *J. Electron Microsc. Tech.* **1989,** *13*, (1) and (2).

[4] I. K. Jordan, C. J. Rossouw, R. Vincent, *Ultramicroscopy* **1991,** *35*, 237.

[5] M. H. Loretto, *Electron Beam Analysis of Materials*, Chapman & Hall, London, **1984**.

[6] J. W. Steeds, *Introduction to Analytical Electron Microscopy*, (Eds.: J. J. Hren, J. I. Goldstein, D. C. Joy, Plenum Press, New York, **1979,** 387.

[7] J. P. Morniroli, J. W. Steeds, *Ultramicroscopy* **1992,** *45*, 219.

[8] J. W. Steeds, J. P. Morniroli, *Reviews in Mineralogy*, **1992,** *27*, 37.

[9] P. M. Jones, G. M. Rackham, J. W. Steeds, *Proc. R. Soc.* **1977,** *A354*, 197.

[10] C. Sung, D. B. Williams, *J. Electron Microsc. Tech.* **1991,** *17*, 95.

[11] See references in [10] and in Sec. 5.4. of [2].

[12] B. F. Buxton, J. A. Eades, J. W. Steeds, G. M. Rackham, *Phil Trans. R. Soc.* **1976,** *A281*, 171.

[13] a) M. Tanaka, *J. Electron. Microsc. Tech.* **1989,** *13*, 27; b) M. Tanaka, *Acta. Cryst.* **1994,** *A50*, 261.

[14] J. W. Steeds, R. Vincent, *J. Appl. Cryst.* **1983,** *16*, 317.

[15] P. B. Hirsch, A. Howie, R. B. Nicholson, D. W. Pashley, M. J. Whelan, *Electron Microscopy of Thin Crystals*, Robert Krieger, Florida, **1977**.

[16] R. C. Ecob, *Scr. Metall.* **1986,** *20*, 1001.

[17] D. M. Bird, *J. Electron. Microsc. Tech.* **1989,** *13*, 77.

[18] J. W. Steeds, P. M. Jones, J. E. Loveluck, K. Cooke, *Phil. Mag.* **1977,** *36*, 309.

[19] J. W. Steeds, *Ultramicroscopy* **1988,** *26*, 1.

[20] H. Matsuhata, J. W. Steeds, *Phil. Mag.* **1987,** *B55*, 17.

[21] J. R. Sellar, D. Imeson, C. J. Humphreys, *Acta. Cryst.* **1980,** *A36*, 686.

[22] B. F. Buxton, *Proc. R. Soc.* **1976,** *A350*, 335.

[23] B. K. Vainshtein, B. B. Zvyagin, A. S. Avilov in *Electron Diffraction Techniques*, Vol. 1 (Ed.: J. M. Cowley), Oxford University Press, Oxford **1992**, Chap. 6.

[24] D. L. Dorset, *Adv. Electron. Electron Phys.* **1994,** *88*, 111.

[25] D. M. Bird, *J. Electron Microsc. Tech.* **1989,** *13*, 77.

[26] R. Vincent, D. M. Bird, J. W. Steeds, *Phil. Mag.* **1984,** *A50*, 745.

[27] D. M. Bird, S. McKernan, J. W. Steeds, *J. Phys. C* **1985,** *18*, 499.

[28] R. Vincent, R. L. Withers, *Phil. Mag. Lett.* **1987,** *56*, 57.

[29] R. Vincent, D. R. Exelby, *Phil. Mag. Lett.* **1991,** *63*, 31.

[30] R. Vincent, D. R. Exelby, *Acta Cryst.* **1995,** *A51*, 801.

[31] R. Vincent, P. A. Midgley, *Ultramicroscopy* **1994,** *53*, 271.

[32] D. M. Bird, M. Saunders, *Ultramicroscopy* **1992,** *45*, 241.

[33] B. C. McCallum, J. M. Rodenburg, *Ultramicroscopy* **1993,** *52*, 85.

[34] W. J. Vine, R. Vincent, P. Spellward, J. W. Steeds, *Ultramicroscopy* **1992,** *41*, 42.

[35] R. Vincent, W. J. Vine, P. A. Midgley, P. Spellward, J. W. Steeds, *Ultramicroscopy* **1993,** *50*, 365.

[36] R. A. Herring, G. Pozzi, T. Tanji, A. Tonomura, *Ultramicroscopy* **1995,** *60*, 153.

[37] D. Cherns, *Inst. Phys. Conf. Ser.* **1991,** *117*, 549

[38] R. Vincent, J. Wang, D. Cherns, S. J. Bailey, A. R. Preston, J. W. Steeds, *Inst. Phys. Conf. Ser.* **1987,** *90*, 233.

[39] I. K. Jordan, A. R. Preston, L. C. Qin, J. W. Steeds, *Inst. Phys. Conf. Ser.* **1989,** *98*, 131.

[40] D. Cherns, R. Touaitia, A. R. Preston, C. J. Rossouw, D. C. Houghton, *Phil. Mag.* **1991,** *A64*, 597.

[41] I. K. Jordan, D. Cherns, N. Grigorieff, M. Hockly, P. C. Spurdens, M. R. Aylett, E. C. Scott, *Inst. Phys. Conf. Ser.* **1991,** *117*, 563.

[42] N. Grigorieff, D. Cherns, M. J. Yates, M. Hockly, S. D. Perrin, M. R. Aylett, *Phil. Mag.* **1993,** *A68*, 121.

[43] D. Cherns, C. J. Kiely, A. R. Preston, *Ultramicroscopy* **1988,** *24*, 355.

[44] Y. Atici, D. Cherns, *Ultramicroscopy* **1995,** *51*, 435.

[45] X-F. Duan, D. Cherns, J. W. Steeds, *Phil. Mag.* **1994,** *A70*, 1091.

[46] D. Cherns and A. R. Preston, *Proc. of 11th Int. Congress on Electron Microsc. Kyoto.* (Eds.: T. Imura, S. Maruse and T. Suzuki) **1986,** *1*, 721.

[47] C. T. Chou, A. R. Preston, J. W. Steeds, *Phil. Mag.* **1992,** *A65*, 863.

[48] D. Cherns, J.-P. Morniroli, *Ultramicroscopy* **1994,** *53*, 167.

[49] J.-P. Morniroli, J. W. Steeds, *Inst. Phys. Conf. Ser.* **1991,** *119*, 417.

[50] M. A. Al-Khafaji, D. Cherns, C. J. Rossouw, R. Hull, *Inst. Phys. Conf. Ser.* **1991,** *119*, 51.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

1.6 Low-Energy Electron Microscopy

1.6.1 Introduction

Low-energy electron microscopy (LEEM) is a surface-imaging method in which the surface is illuminated at normal incidence by slow electrons and the elastically reflected electrons are used to form an image of the surface. It is based on the high back-scattering cross-section which all materials have at low energies. In particular, in crystalline materials the back-scattered electrons are concentrated in one or more diffraction beams. One of them, usually the specular beam [the (00) beam], can be used to image the surface, wholly analogous to the manner in which the image is generated in the transmission electron microscope by the transmitted or by diffracted beams. LEEM is particularly useful for the study of processes on metal and semiconductor single-crystal surfaces. It is easily combined with other surface-imaging methods such as mirror electron microscopy (MEM) or photoemission electron microscopy (PEEM; see Chap. 6, Sec. 1 of this Volume). The lateral resolution of LEEM is limited by the aberrations of the cathode objective lens, whose cathode is the specimen, to the 1–10 nm range. The field of view can be as large as 100 μm, depending upon the lens used, but the depth resolution is in favorable cases in the subangstrom range. Many surface science and technology problems do not require a better resolution but a large field of view. Therefore, LEEM has a wide range of applications.

1.6.2 Theoretical Foundations

LEEM can be best understood by looking at the low-energy electron diffraction (LEED) pattern which is formed at the back focal plane of the cathode lens. Even in the absence of lens aberrations the LEED pattern is not the Fourier transform of the potential distribution in the surface layer of the specimen with which the incident electron wave interacts. This is due to the strong electron–specimen interaction which makes the first Born approximation invalid. As a consequence, the kinematical diffraction theory is insufficient for the calculation of the diffraction beams. Rather, the dynamical theory has to be used as described in various textbooks [1–4].

However, presently no quantitative information is drawn from intensity differences in LEEM images. Only the magnitude of the difference which determines the contrast is important. As an example,

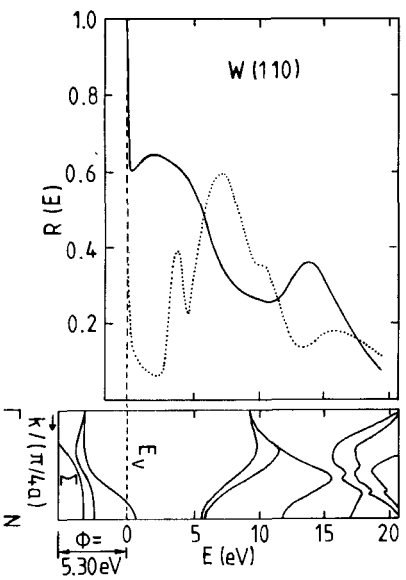


Figure 1. Specular reflection coefficient of the W(110) and W(100) surfaces as a function of energy (top) and band structure of tungsten along the [110] direction (Σ direction) [6].

Fig. 1 shows the intensity $I_{00}(E)$ of the specularly reflected beam from a W(110) and a W(100) surface as a function of energy E [5, 6]. The two I_{00} curves differ drastically in their energy dependence, so that it is always possible to find strong contrast between (110) and (100) oriented grains, for example at about 2 or 14 eV. On the other hand, no contrast may be seen if the energy is chosen wrongly, for example at about 6 or 11 eV. The figure also shows that reflectivities as high as 60% may be obtained. Such high values are connected with band gaps or regions with a low density of states (steep bands) in the band structure of the crystal along the propagation direction of the wave. For energies in band gaps no wave propagation is possible, and the wave is strongly damped and totally reflected: the crystal is a reactive medium. Ideally, the reflectivity should be 100%, but inelastic scattering, which is predominantly in the forward direction, and phonon and defect scattering decrease it to the observed value. The missing electrons appear in the specimen current [6]. Similar considerations apply to energy ranges with a low density of states, but with increasing energy, part of the intensity is lost into other diffraction channels. For example, at 13.6 eV where $I_{00}(E)$ has a maximum again on the W(110) surface, 10 diffraction beams in addition to the specular beam are excited as easily seen by an Ewald construction (taking the inner potential $V_0 = 17.4\,\text{eV}$ into account). In addition, elastic back scattering decreases and inelastic scattering increases with energy and reduces I_{00} even more.

Diffraction contrast is not only strong between grains with different orientations but also between regions on a single-crystal surface which differ in surface structure, such as between reconstructed and unreconstructed surface regions or clean and adsorbate-covered regions, because these regions also differ in their $I_{00}(E)$ dependence. Examples which will be discussed later are the contrast between (1×1) and (7×7) structures on Si(111) or between clean Mo(110) and copper monolayer islands on this surface. A frequent situation in surface science is the presence of an ordered overlayer in two or more azimuthal orientations. They are indistinguishable at normal incidence because the diffraction conditions are identical. Contrast can be obtained by changing the diffraction conditions, that is, by tilting the beam, or by imaging with one of the nonspecular overlayer beams. The resulting contrast has been termed tilted bright field and dark field contrast, respectively, while the contrast described previously has been called bright field contrast.

The contrast discussed up to now depends upon the periodic arrangement of the atoms in the surface region of the specimen. One of the main goals of surface microscopy is, however, to image deviations from this periodic arrangement, that is, surface defects. The most important defects are monoatomic steps, because many surface properties and processes depend strongly upon them. LEED is uniquely suited for the determination of the average step distribution which determines the profiles of the diffraction spots [7, 8]. The profiles are analyzed in terms of the kinematic theory because the backscattered intensity may be separated into a dynamic and a kinematic term, $|F|^2$ and $|G|^2$, respectively. The structure factor $|F|^2$ contains the information on the size and shape of the two-dimensional unit cell and the atoms in it, the lattice factor $|G|^2$ the size and shape of the periodic arrangement of the unit cells. A step terminates this arrangement by introducing a vertical phase shift which changes $|G|^2$. In bright field LEEM the step contrast can be easily understood in terms of this phase shift. A step of height d produces a path difference $2d$ between the waves reflected from the adjoining terraces. When $(2n+1)\lambda/2 = 2d$, the waves are out of phase and interfere destructively upon defocusing or as a consequence of the limited phase contrast transfer function of the objective lens. By determining the λ values—that is, the energies $E = 1.5/\lambda^2$ (in electronvolts when λ is in nanometers)—at which the contrast is strongest, the step height can be determined with high accuracy. LEEM thus allows the determination of the local step distribution. An example is shown in Fig. 2.

Another contrast which is also a consequence of path differences between

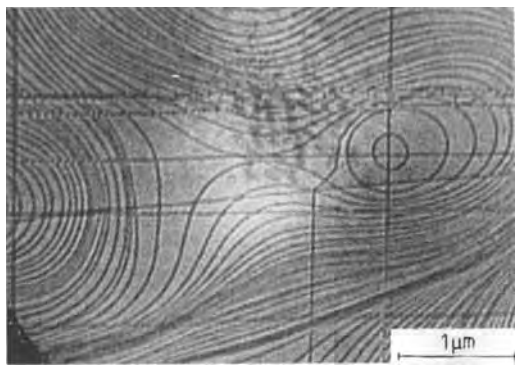


Figure 2. LEEM image of the step distribution on an Mo(110) surface. The rounded steps were formed by sublimation during the preceding high-temperature treatment, while the straight steps are the result of glide during cooling. The electron energy is 14 eV [9].

different areas on the specimen is the quantum size effect contrast which appears in thin films with parallel top and bottom boundaries. It is the electron–optical counterpart to reflection-reducing or -enhancing layers in optics: the reflectivity has a maximum when $2d = n\lambda_f$ and a minimum when $2d = (2n+1)\lambda_f/2$ (λ_f is the wavelength in the film). Thus, the thickness, and thickness differences, can be determined with high accuracy. Finally, it has to be kept in mind that the specimen is the cathode in a cathode lens. This means that the specimen is in a high field, and any deviation from planarity will cause a field distortion which will deflect the electrons. These deflections produce a topographic contrast which is frequently very helpful. However, if the field distortions of the surface features overlap, image interpretation becomes difficult if not impossible.

An important aspect of any surface microscopy is the information depth. In LEEM, as in any other surface probe which uses slow electrons, the information depth is, in general, determined mainly by

inelastic scattering but also by thermal diffuse scattering and elastic back scattering. Elastic back scattering is important mainly at very low energies at which, frequently, only the specular beam can be excited, as discussed above, and a band gap exists. The information depth is then of the order of the wavelength of the electron, which is 3–4 Å at zero energy for realistic mean inner potentials due to the exponential damping in total reflection. Thermal diffuse scattering, in particular multiphonon scattering which causes scattering predominantly far from the (0, 0) beam, becomes important at high temperatures.

Inelastic scattering due to collective and one-electron excitations dominates the information depth at most temperatures. The inelastic mean free path has a minimum value of 3–5 Å around 50 eV which is predominantly caused by plasmon excitation. Toward lower energies—below the plasmon excitation threshold—it increases rapidly, and toward higher energies more slowly. The attenuation length in the LEEM energy range (0 eV to about 200 eV) is, thus, ranges from a few angstroms to 10 Å. In materials with fully occupied d bands, higher values are possible at very low energies because the possibilities of low-energy excitations are rather limited in them. If the d band is only partially occupied, low-energy excitations with high cross-section are possible, and the inelastic mean free path may decrease to less than 3 Å (see this Chap. 4, Sec. 1.1.6 of this Volume). On the other hand, in large band gap materials much larger values may occur. Depth information may also be obtained indirectly, for example on buried interfaces via the strain field of the interfacial dislocation, which may give sufficient diffraction contrast at layer thicknesses of the order of 100 Å.

Additional information on electron–specimen interaction and on contrast formation may be found in the literature [10–12].

1.6.3 Instrumentation

A LEEM instrument differs in several aspects from conventional electron microscopes. The very nature of its purpose, the imaging of well-defined surfaces, requires an ultrahigh vacuum (UHV) with a base pressure in the low 10^{-10} to high 10^{-11} mbar range. Similarly to the mirror microscope, illuminating and imaging must be separated if an image is to be produced. Otherwise only diffraction is possible. Finally, the specimen must be at a high negative potential or the complete optical system must be at high positive potential if the specimen should be at ground potential. Otherwise, an LEEM instrument makes use of the usual components of standard electron microscopes, albeit using UHV technology. The first (ill-fated) system was built with glass technology using electrostatic lenses and 90° deflection. Until recently, all subsequent systems were all-metal systems with magnetic lenses and 60° deflection [9, 13, 14]. The desire to have the specimen at ground potential has led to the development of electrostatic lens systems, the first of which made use of double 45° deflections [15, 16]. More recently, instruments with 90° deflection have been proposed again [17, 18].

As an example, the most sophisticated instrument at present will be described

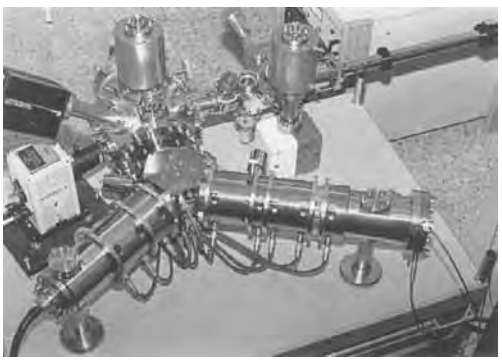


Figure 3. Overall view of a versatile LEEM instrument [12].

[13, 19]. Figure 3 shows its overall view, and Fig. 4 its cross-section, with an energy filter added. The design concept of the instrument is based on the desire to use it not only for LEEM but also for Auger electron emission microscopy (AEEM) and secondary electron emission microscopy (SEEM). In these two imaging modes the electron energies in the illumination and in the imaging sections are different: the illumination system must allow a wide range of illumination conditions and the imaging section must be able to accommodate an imaging energy filter. For efficient SEEM, primary beam energies up to about 500 eV are needed because the secondary electron yields of most materials peak at a few hundred electronvolts, while the secondary electron energy distribution peaks between a few electronvolts and 10 eV. In AEEM, primary electrons up to a few kiloelectronvolts are necessary in

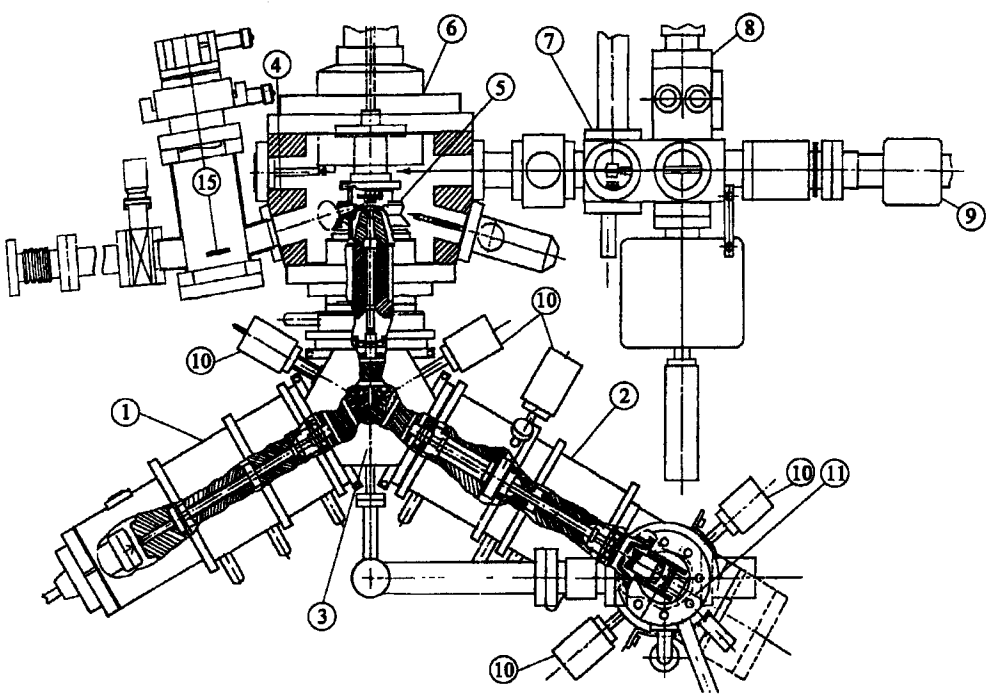


Figure 4. Cross-section of the instrument shown in Fig. 3 but with an energy analyzer added: 1, illumination column; 2, imaging column; 3, beam separator; 4, specimen chamber; 5, cathode lens; 6, specimen manipulator; 7, specimen preparation chamber; 8, air lock; 9, transfer rod; 10, aperture manipulators; 11, energy filter; post-filter projective and detection system; 15, refocusing mirror chamber for synchrotron radiation.

order to be able to detect all elements, while the characteristic energies of the Auger electrons range from a few tens of electronvolts to about 1 keV.

For LEEM, and in particular for LEED, the specimen should be illuminated with a parallel beam; intensity is of minor concern because of the strong elastic back scattering. For AEEM and SEEM, parallel illumination is unimportant, but a high illumination intensity is required, at least for AEEM, because the Auger electron signals are weak and ride on a strong background. Without high intensity, excessive image acquisition times would be necessary for images with a sufficient signal-to-noise ratio. In order to achieve this flexibility an LaB_6 cathode is used plus a double condensor and a transfer lens which images the demagnified cross-over either onto the back focal plane of the objective lens for LEEM and LEED or—together with the objective lens—onto the specimen for AEEM and SEEM. In the LEEM and LEED modes the beam aperture angle in the back focal plane and, therefore, the illuminated area on the specimen is controlled by an aperture in the beam separator. The beam separator, which is designed to be double focusing, consists of several sections which can be excited independently of each other so that primary and secondary electrons with different energies can be deflected by the same angle (60°) [20]. The 'close-packed prism array' used in the instrument in Figs. 3 and 4 is schematically shown in Fig. 5a with the central ray path indicated. The objective lens is a magnetic triode lens, that is, a magnetic lens in which the pole piece next to the specimen is electrically isolated (Fig. 5b) so that its potential can be varied from ground potential to near-

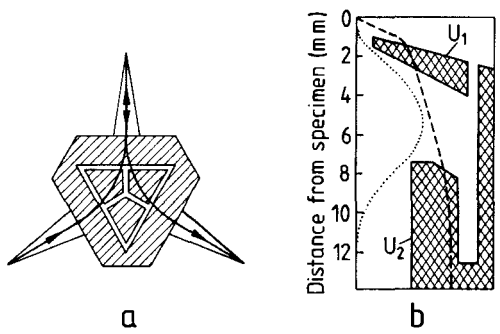


Figure 5. (a) Double-focusing separator for incident and emitted beams with different energies. (b) Magnetic triode cathode objective lens. The axial electrostatic potential and magnetic field are indicated by dashed and dotted lines, respectively.

specimen potential. This allows the study of surfaces at low fields. Other cathode lenses can and have been used as well [12]. The contrast aperture is not placed in the back focal plane of the objective lens—where the LEED pattern is located—as is usually done in standard electron microscopes—for two reasons: (1) for dark field imaging, that is, imaging with nonspecular beams, the aperture has to be shifted laterally and would intercept the incident beam for large shifts, and (2) the aperture would limit the intensity in the AEEM and SEEM illumination modes. Therefore, the contrast aperture is placed further downstream in an image of the LEED pattern. The objective lens is focused in such manner that the image of the specimen is in the center of the separator, where it is, in principle, achromatic. For selected-area LEED, a field-limiting aperture can be placed in this position. However, in most applications the illumination aperture is sufficient for this purpose, and the port for the manipulator of the field limiting aperture is used for a fluorescent screen with which the primary beam can be inspected (cathode

inspection, illumination system alignment). The contrast aperture in the image plane of the LEED pattern limits not only the angular aperture used in further image formation but acts also to a certain extent as a rough energy filter due to the (small) chromatic aberration of the deflection field.

The imaging system after the deflection field must allow the placement of either the image or the LEED pattern at the input side of the energy filter. For proper matching and magnification selection, several lenses are necessary [13, 19]. The energy filter is a 90° electrostatic sector in which the electrons are decelerated to about 1/10th of the energy which they have in the rest of the instrument except in the cathode field region. The final energy-filtered image is obtained by a double-gap projective lens on a double-channel plate–fluorescent screen detector. Of course, if the instrument is to be used only for LEEM, LEED, or mirror and emission electron microscopy with very slow electrons such as thermionic electron emission microscopy (TEEM) or near-ultraviolet PEEM, no energy filter, fewer lenses and a simpler beam separator are needed, and the instrument can be much simpler, as is the case for other instruments [9, 21]. In addition to the optical components mentioned up to now, the instrument contains a number of deflection systems for beam alignment and three stigmators, one in the objective lens, and one each in the illumination and imaging columns. The last two not only correct lens astigmatism but also the residual cylinder lens action of the beam separator.

The objective lens is located in a large specimen chamber with six ports pointing at the specimen position for mounting accessories for in situ experiments (gas sources, evaporators, etc.) and for illumination with a high-pressure mercury short arc lamp for PEEM. The specimen is mounted on an eccentric *x-y-z* manipulator on which it can be heated to 1600°C. Specimen exchange is through an airlock via a specimen preparation chamber. The system is pumped by sputter ion and titanium sublimation pumps, and has a base pressure in the 10^{-11} mbar range in the specimen chamber after a bake-out at 180°C. LEEM image acquisition is usually via a video camera, but a still camera can be used for high image quality. Spectroscopic images are recorded with a slow scan Peltier-cooled charge-coupled device camera. Figure 6a shows an electron beam-excited AEEM image, and Fig. 6c the local photoelectron spectrum which was obtained from a series of synchrotron radiation-excited photoelectron images taken in small energy increments (0.5 eV) and with good energy resolution (0.5 eV) such as the one shown in Fig. 6b.

Provided that vibrations, current and high-voltage fluctuations, and alternating current fields are low enough, the resolution of an LEEM is limited mainly by the lens aberrations of the cathode lens, the aberrations of the beam separator and—in a spectroscopic instrument—the aberrations of the energy filter. With a well-designed beam separator and energy filter and a well-aligned system the cathode lens determines the ultimate resolution, in particular its acceleration field region. In an idealized situation the lens can be divided into two sections: a homogeneous acceleration field, which forms a virtual image behind the specimen, and a subsequent conventional electron lens, which forms a real image of the virtual image. The

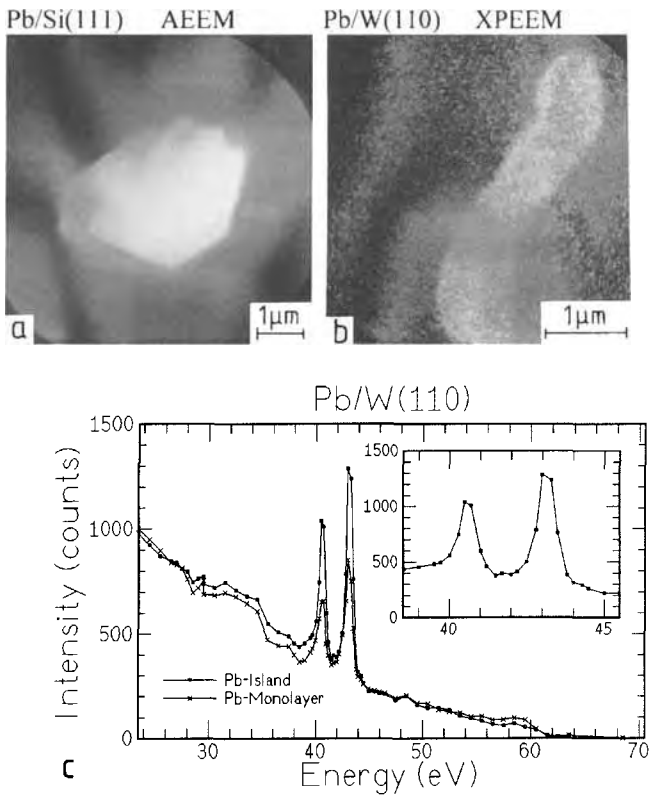


Figure 6. Specimen characterization in a spectroscopic LEEM. Lead layer on (a) an Si(111) surface and (b, c) a W(110) surface. (a) Electron-excited AEEM image of a lead crystal (primary energy 2.5 keV, imaging energy 94 eV, 5 s exposure); (b) synchrotron radiation-excited XPEEM image; (c) X-ray ($h\nu = 60.7$ eV) photoelectron spectrum of a 0.4 μm^2 area on a lead crystal, and of a 0.8 μm^2 area of a lead monolayer obtained from a sequence of images (b) (20 s/image) [23].

aberrations of the unavoidable acceleration field are dominating. They can be easily calculated analytically. For a fixed aperture radius r_A at the exit of the homogeneous field the spherical and chromatic aberrations are proportional to $\Delta E/F$ and E/F, respectively, where F is the field strength at the cathode, E the average start energy of the electrons at the specimen, and ΔE the energy width [22]. The actual values for a specific example are shown in Fig. 7a. For comparison, Fig. 7b shows the dependence of the resolution upon the start energy for three cathode lens types used in LEEM instruments for an energy width of 0.5 eV [24]. Reduction of the energy width to $\Delta E = 0.1$ eV reduces the resolution δ somewhat, but $\delta = 4$ nm is about the limit at a start energy of 2 eV. A further improvement of δ is possible with elaborate correction elements [17] but has not been realized yet.

1.6.4 Areas of Application

1.6.4.1 Clean Surfaces

Because of its high depth resolution, LEEM is very well suited for the study of surface topography. The imaging of monoatomic steps has already been mentioned (Fig. 2), but etch pits, hillocks, or screw dislocations can be imaged as well. The real-time and high-temperature capabilities of LEEM also allow the life study

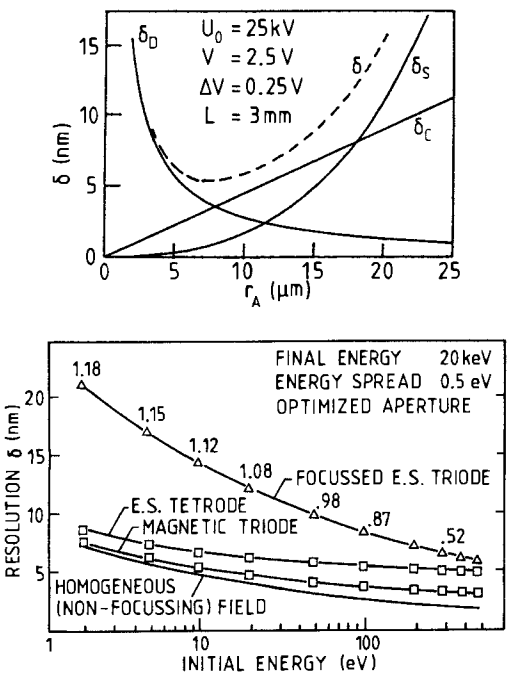


Figure 7. Resolution of the cathode lens. (a) Dependence of spherical aberration resolution δ_S, chromatic aberration resolution δ_C and diffraction disk δ_D resolution upon aperture radius r_A of a homogeneous field of length L [22]. (b) Resolution of the electrostatic triode, electrostatic tetrode, and magnetic triode (Fig. 6) as a function of start energy. The values on the upper curve are the field strengths at the specimen [24].

of changes in topography due to surface diffusion and other processes such as sublimation. For the same reasons, the microstructural aspects of structural phase transitions are particularly attractive for LEEM. A well-known example is the Si(111)–(1 × 1) ↔ (7 × 7) phase transition, which can occur in a wide variety of morphologies with little difference visible in laterally averaging techniques such as LEED (Fig. 8) [25]. Domain growth on reconstructed surfaces [26a] or facetting of vicinal surfaces [26b] are other problems well-suited for LEEM.

1.6.4.2 Adsorption Layers

Many foreign atoms or molecules on surfaces have overall attractive lateral interactions even if the nearest-neighbor interactions are repulsive. Therefore, they form islands which produce structural contrast. This is also the case when the adsorbate and substrate have the same lateral periodicity as in submonolayers of copper on Mo(110). The contrast between island and substrate is due to differences in back-scattering cross-section and phase. In general, however, the adsorbate layer has a lateral periodicity different from that of the substrate and, as a consequence, forms a number of equivalent domains. Dark field imaging with superstructure reflections characteristic for a certain domain then allows them to be imaged individually. An example is shown in Fig. 9. The two (5 × 1) domains which appear dark in the bright field image (Fig. 9b) are easily distinguished by imaging with their corresponding (1/5, 0) order beams. The correlation between domain boundaries and steps is also visible.

Most adsorbates experience order–disorder transitions with increasing temperature before desorption occurs. Many of these transitions appear to be continuous in laterally averaging studies, for example with LEED. Gold adsorbates on Si(111) are again a good example [27]. Another is the misfit flip phase transition in the copper double layer on Mo(110) [28]. This transition layer is also a demonstration of the usefulness of being able to image at high temperatures, which allows one to follow the kinetics of the strain-induced surface restructuring of the original flat surface into the structure shown in Fig. 10 [29] and the subsequent

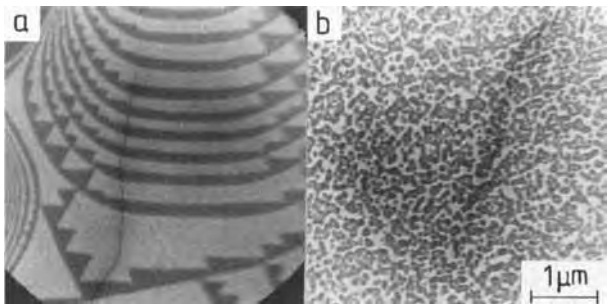


Figure 8. LEEM images taken during the Si(111)–(1 × 1) ↔ (7 × 7) phase transition. Both surfaces show a good (7 × 7) LEED pattern below the transition temperature (electron energy 11 eV) [25].

straightening of the steps during copper desorption. Desorption from the two-phase region (coexisting two-dimensional islands and gas) can also be easily distinguished by LEEM from desorption from the single-phase region (two-dimensional gas only). By slowly lowering the temperature, the condensation of two-dimensional islands from the two-dimensional gas can be studied. An example is shown in Fig. 11 [30]. During cooling, the gold atoms did not diffuse to the steps and condense there but formed numerous strongly elongated islands with one of the two crystallographically equivalent orientations extending over large areas, indicating island–island interactions through the adsorbate-induced strain field in the substrate.

If an adsorbate is deposited locally, then surface diffusion can be studied by following the propagation of the boundary between characteristic adsorbate structures which differ in coverage. Similarly, reaction fronts between two different

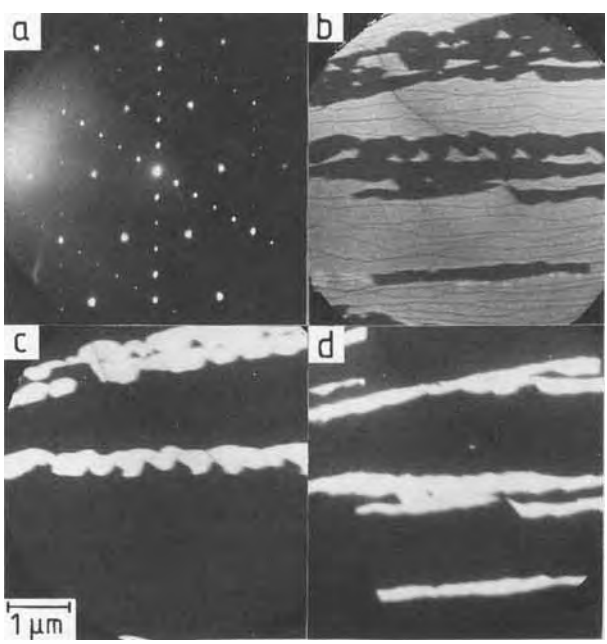


Figure 9. (a) LEED pattern (30 eV), (b) bright field image and (c, d) two dark field images taken with different 1/5 order reflections of a mixture of the $(\sqrt{3} \times \sqrt{3})$ R30° and the (5 × 1) structure on Si(111) in two domain orientations (electron energy 6 eV) [27].

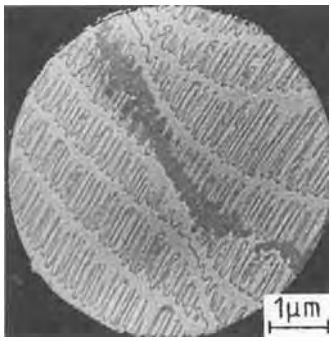


Figure 10. Step break-up of an Mo(110) surface covered with a copper double layer at about 900 K (electron energy 3 eV) [28].

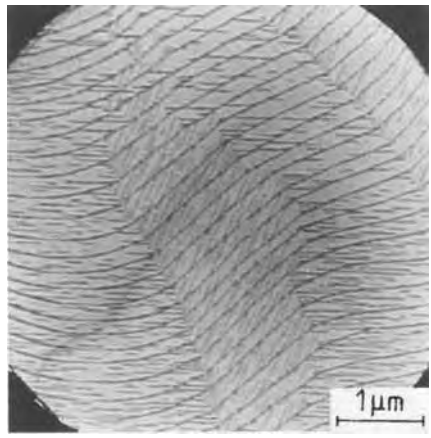


Figure 11. Two-dimensional condensation of gold from the two-dimensional gas phase on Mo(110) (electron energy 14 eV) [29].

adsorbates can be used to follow chemical reactions on surfaces [31]. This can be done also by MEM [32] and PEEM [33] if characteristic work function differences exist between the reacting adsorbates, a situation frequently encountered such as in the case of co-adsorbed oxygen and carbon monoxide.

Adsorption of impurities from the interior of the crystal, that is, segregation, is also a good subject for LEEM. This is illustrated in Fig. 12 by the segregation of carbon to the (110) surface of molybdenum, where it forms Mo_2C crystals in a large morphological variety, depending upon the carbon concentration, cooling rate, and other parameters. Figure 12 also shows the influence of the electron energy on the contrast: at 3 eV the Mo_2C-covered regions are dark and the clean molybdenum surface bright, while at 9 eV the reverse is true.

Another frequent surface phenomenon is the formation of two-dimensional compounds by interactions with reactive gases. The early stages of oxidation have been studied with laterally averaging techniques for decades, but little is known about the processes on a microscopic level. Although not much work has been done with LEEM, a good deal of insight has already

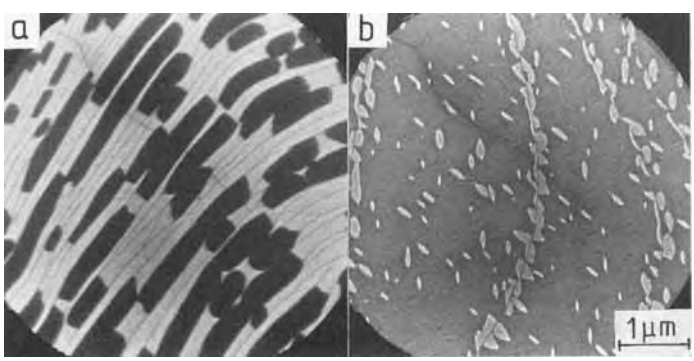


Figure 12. Mo_2C formation on Mo(110) by carbon segregation from the bulk: electron energies of (a) 3 eV and (b) 9 eV [29].

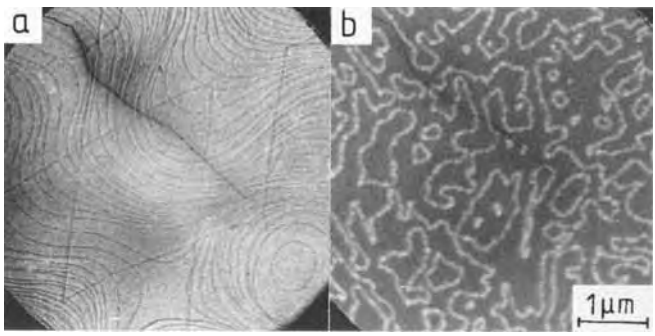


Figure 13. Clean W(100) surface ((a) electron energy 7 eV) and onset of two-dimensional oxide formation ((b) electron energy 16 eV) [34].

been gained. While in many surface processes, steps play an important role, this does not seem to be the case in the high-temperature oxidation of W(100) at low pressure. Once domains of chemisorbed oxygen have formed—which involves a considerable amount of surface rearrangement—the domain boundaries are the weak points that determine the further oxidation process. Figure 13 shows the initial step structure on a clean W(100) after a high-temperature flash and the transition from the chemisorbed layer to the two-dimensional oxide which starts at the domain walls and spreads from them rapidly across the surface [34].

Two-dimensional silicon nitride formation in the early stages of the nitridation of Si(111) in ammonia at high temperatures proceeds quite differently. The nucleation clearly occurs at steps, and the (8×8) nitride structure then spreads across the terraces in a manner similar to the growth of the (7×7) structure on the clean surface. Again, depending upon nitridation conditions and the purity of the surface, a wide variety of morphologies and kinetics is observed. A typical frame from a video taken during this process is shown in Fig. 14. The combination with scanning tunneling microscopy has proven very useful in understanding structure and growth of the two-dimensional nitride [35].

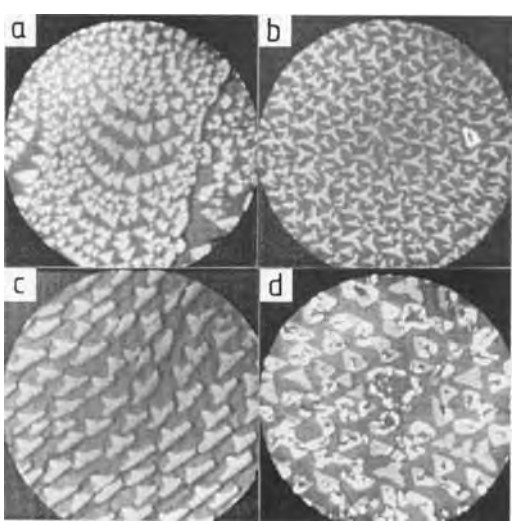


Figure 14. Nucleation and growth of two-dimensional nitride with an (8×8) structure on Si(111) by reaction with ammonia at 1150 K (electron energy 3 eV). Video frames are from growth experiments [35].

1.6.4.3 Thin Films

One of the major application areas of LEEM is the study of the early stages of the growth of thin films, in particular of how surface imperfections influence the growth and how reaction layers grow.

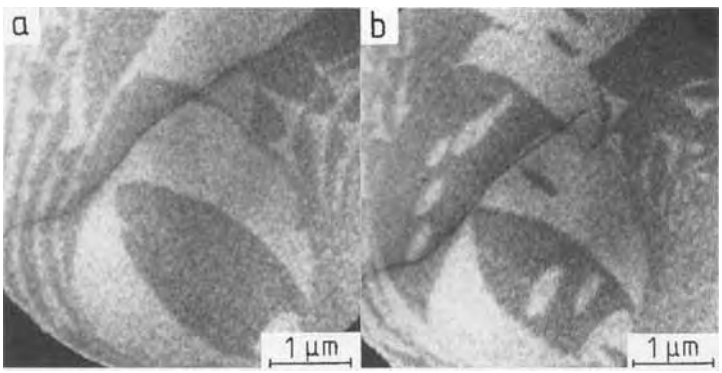


Figure 15. Homoepitaxy of silicon on Si(100) at (a) very low supersaturation and (b) low supersaturation (electron energy 5 eV). Video frames are from a growth experiment [36].

Most metal–metal and metal–semiconductor film–substrate combinations alloy or react, at least at high temperatures, but there are also some less complicated systems, the simplest being homoepitaxy. One of the most important processes is the homoepitaxy of silicon on Si(100), which has also been studied with LEEM. At low supersaturation, at which no nucleation can take place on the terraces, new terraces nucleate only at imperfections, and spread over the existing terraces in a step–flow growth mode. The growth rate is highly anisotropic, reflecting the large difference between the step energies of steps parallel and perpendicular to the dimer rows on this (2×1)-reconstructed surface. This leads to the almond-shaped terraces seen in Fig. 15a. At somewhat higher supersaturation, nucleation on the terraces occurs (Fig. 15b), and the images rapidly become too complex for analysis [36].

The next simplest growth process is heteroepitaxy without alloying or reaction. Typical examples are copper, silver, and gold on W(110). Below 400–500 K these layers grow in a quasi-monolayer-by-monolayer mode, but at the upper temperature limit they grow at significantly different rates on different terraces. Some terraces have preferred nucleation sites from which the monolayers grow rapidly along the terraces, obviously incorporating atoms that diffuse across the steps from other terraces. This produces considerable thickness variations from terrace to terrace, which can be studied well by making use of the quantum size effect (Fig. 16) [37a]. At high temperatures, only one or two adsorbed layers are formed on which flat three-dimensional crystals grow. An example can be found in the chapter on SPLEEM (see Chap. 5, Sec. 3 of this Volume), in which also the agglomeration of low temperature-deposited layers upon annealing are discussed as well as alloying in miscible film–substrate systems.

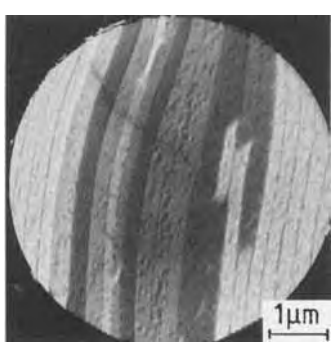


Figure 16. Terrace-limited monolayer-by-monolayer growth of copper on Mo(110) imaged via quantum size effect contrast (electron energy 4 eV) [37a].

The most complex film growth process involves reaction with the substrate. The growth and disintegration of silicide layers on silicon is an important example of this type. $CoSi_2$ layers have attracted considerable attention because of their possible applications in microelectronics. An LEEM study showed that these layers tend to grow as three-dimensional crystals with a clean silicon surface in between. The $CoSi_2$ crystals thus act as scavengers for cobalt atoms. Heating such a layer to a high temperature causes it to break up into many small three-dimensional crystals, from which a few large ones form via Oswald ripening at very high temperatures. $CoSi_2$ sublimes more slowly than silicon and, therefore, large hillocks grow, with $CoSi_2$ crystals on top (Fig. 17). This example shows that three-dimensional structures can also be imaged quite well, in spite of the field distortion caused by them, provided they are well separated. An additional condition for seeing structure in LEEM is that the faces of the three-dimensional crystal may not be tilted very much. This is illustrated in Fig. 18 with a Cu_3Si crystal obtained by copper deposition on Si(111). The structure of the crystal is identified by selected-area LEED, which simultaneously shows that the crystal has tilted top faces. The specular beams from these faces are off-axis and, therefore, are intercepted by the contrast aperture so that the crystal appears dark [37b]. In contrast to $CoSi_2$ crystals, the Cu_3Si crystal is not surrounded by clean silicon but by a two-dimensional copper silicide, which causes strong step faceting and

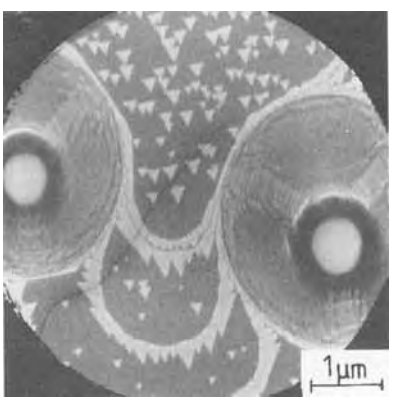


Figure 17. Hillocks on Si(111) with $CoSi_2$ crystals on top, formed by faster sublimation of the substrate (electron energy 10 eV) [25].

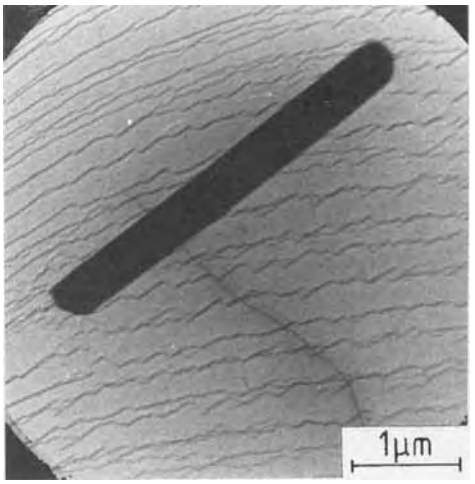


Figure 18. Cu_3Si crystal surrounded by a two-dimensional copper silicide layer on Si(111) (electron energy 5 eV) [37b].

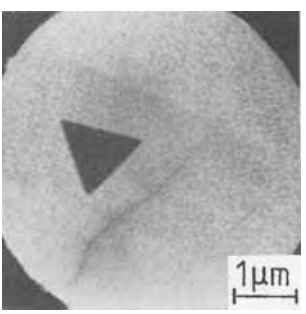


Figure 19. Au–Si eutectic 'particle' on an Si(111) surface at about 900 K [27].

agglomeration. Not all particles which are dark at all energies in LEEM are faceted. The dark triangle in Fig. 19 is not a crystal but a liquid Au–Si eutectic on an Si(111) surface which has well-defined equilibrium boundaries. It is dark because there is no well-defined specular beam due to the lack of crystallinity [27].

1.6.5 Discussion

When should LEEM be used for surface or thin film studies, and when not? The situations when it should not be used are obvious: when a routine resolution of better than 10 nm is needed, when the specimen is not UHV compatible because of high vapor or dissociation pressures, when it is very electron beam-sensitive due to electron-stimulated dissociation or desorption, or when it is very insulating or very rough. Another limitation is the gas pressure at the specimen. These situations are, however, not as LEEM-exclusive as they appear. Further improvements in objective and beam separators should allow a skilled operator to reach the middle 1 nm range for suitable objects. Specimen cooling will reduce the pressure problems, and the use of electron energies below the thresholds for dissociation or desorption, which are easily accessible in LEEM, will eliminate specimen damage. Charging of insulating samples can be frequently eliminated by heating, sometimes by choosing electron energies at which the secondary electron emission coefficient is unity. The only remedy for roughness is smoothing the surface by proper preparation procedures. As far as admissible pressures are concerned, LEEM images have been taken at pressures up to the 10^{-6} mbar range without noticeable image deterioration. This is possible because beam separator, illumination, and imaging columns are pumped separately from the specimen chamber, and the narrow constrictions between the chamber and the columns allow the maintenance of a large pressure gradient. It is mainly high-voltage stability which limits the maximum operating pressure. A major prerequisite is, of course, that the system is always clean, otherwise carbon contamination will make well-defined surface studies impossible.

The major competitors to LEEM in surface studies are high-energy reflection electron microscopy (REM; see Chap. 4, Sec. 1.2 of this Volume) and standard scanning electron microscopy (SEM; Chap. 4, Sec. 2 of this Volume). Both types of instruments are available with surface science-compatible pressures, at least in the specimen region, and bake-out possibilities. Advantages of REM are higher resolution perpendicular to the direction of the electron beam and access to the specimen at nearly normal incidence. A great disadvantage is the strong foreshortening in the beam direction caused by the grazing incidence of the electron beam. The major advantages of SEM are that it can be used for rough surfaces, that local spectroscopy is much faster in a scanning electron microscope equipped with an energy analyzer than in the corresponding LEEM instrument. However, image acquisition in an LEEM is much faster than in an SEM because of the parallel, versus sequential, detection. Also, the surface sensitivity is usually poorer in SEM than in LEEM and the

contrast is more limited. The last two disadvantages can be reduced by going to very low energies, for example by decelerating the electrons immediately before the specimen (scanning LEEM, SLEEM [38, 39]), but this introduces one of the main disadvantages of LEEM, the field distortions on rough samples. A great advantage of LEEM compared to the other two techniques is its easy combination with other imaging modes such as TEEM, PEEM, or SEEM and, in particular, MEM. MEM is very useful on non-crystalline surfaces, for example before the removal of amorphous surface layers such as SiO_2 on crystalline silicon or the (chemo)mechanical polishing layer on single-crystal surfaces. LEED, which is easily obtained by removing the contrast aperture and changing the excitation of the intermediate lens, is indispensable for chemical identification via the diffraction pattern.

Finally, a comparison with the various forms of scanning probe microscopies such as scanning tunneling microscopy (STM) or atomic force microscopy (AFM) is necessary. STM and AFM have an intrinsically higher resolution but are much slower due to their mechanical scanning image acquisition. When operated at high resolution, their field of view is small and vice versa. Operation over a wide temperature range is still a problem, but present STM and AFM instrument developments show that this can be overcome. Another, albeit minor, restriction is in deposition studies, because of the shadowing by the scanning tip. Experience has shown that STM and LEEM are more complementary than competing techniques [35, 40].

1.6.6 Concluding Remarks

LEEM is still a relatively young technique. Only a few instruments presently exist, and the number of studies and publications is still limited, as the latest review [12] shows. For more recent work, see Michely et al. [41], Tromp and Michely [42] and Altman and Bauer [43].

1.6.7 References

[1] J. B. Pendry, *Low Energy Electron Diffraction*, Academic Press, New York **1974**.
[2] M. A. Van Hove, S. Y. Tong, *Surface Crystallography by LEED*, Springer, Berlin **1979**.
[3] L. J. Clarke, *Surface Crystallography*, Wiley, Chichester 1985.
[4] M. A. Van Hove, W. H. Weinberg, C.-M. Chen, *Low-Energy Electron Diffraction*, Springer, Berlin **1986**.
[5] H.-J. Herlt, R. Feder, G. Meister, E. Bauer, *Solid. State Commun.* **1981,** *38*, 973.
[6] H.-J. Herlt, Ph.D. thesis, TU Clausthal, **1982**.
[7] M. Henzler, *Appl. Surf. Sci.* **1982**, *11/12*, 450; M. Henzler, *Appl. Phys.* **1984,** *A34*, 205.
[8] H.-N. Yang, G.-C. Wang, T.-M. Lu, *Diffraction from Rough Surfaces and Dynamic Growth Fronts*, World Scientific, Singapore **1993**.
[9] W. Telieps, E. Bauer, *Ultramicroscopy* **1985,** *17*, 57.
[10] E. Bauer, W. Telieps in *Surface and Interface Characterization by Electron Optical Methods* (Eds.: A. Howie, U. Valdre), Plenum, New York **1988,** p. 195.
[11] E. Bauer in *Chemistry and Physics of Solid Surfaces VII* (Eds.: R. Vanselow, R. Howe), Springer, Berlin **1990,** Chap. 12.
[12] E. Bauer, *Rep. Prog. Phys.* **1994,** *57*, 895.
[13] L. H. Veneklasen, *Ultramicroscopy* **1991,** *36*, 76.
[14] E. Bauer, *Surf. Sci.* **1994,** *299/300*, 102.
[15] K. Grzelakowski, T. Duden, E. Bauer, H. Poppa, S. Chiang, *IEEE Trans. Magn.* **1994,** *30*, 4500.
[16] K. Grzelakowski, E. Bauer, *Rev. Sci. Instrum.* **1996,** *67*, 742.
[17] H. Rose, D. Preikszas, *Optik* **1992**, *92*, 31.
[18] R. M. Tromp and M. Mankos, personal communication.

[19] L. H. Veneklasen, *Rev. Sci. Instrum.* **1992**, *63*, 5513.
[20] V. Kolarik, M. Mankos, L. Veneklasen, *Optik* **1991**, *87*, 1.
[21] R. M. Tromp, M. C. Reuter, *Ultramicroscopy* **1991**, *36*, 91.
[22] E. Bauer, *Ultramicroscopy* **1985**, *17*, 51.
[23] C. Koziol, T. Schmidt, G. Lilienkamp, E. Bauer, to be published.
[24] J. Chmelik, L. Veneklasen, G. Marx, *Optik* **1989**, *83*, 155.
[25] E. Bauer, M. Mundschau, W. Swiech, W. Telieps, *J. Vac. Sci. Technol.* **1991**, *A9*, 1007.
[26] (a) R. J. Phaneuf, N. C. Bartelt, E. D. Williams, W. Swiech, E. Bauer, *Surf. Sci.* **1992**, *268*, 227; (b) R. J. Phaneuf, N. C. Bartelt, E. D. Williams, W. Swiech, E. Bauer, *Phys. Rev. Lett.* **1993**, *71*, 2284.
[27] W. Swiech, E. Bauer, M. Mundschau, *Surf. Sci.* **1991**, *253*, 283.
[28] E. Bauer, *Appl. Surf. Sci.* **1992**, *60/61*, 350.
[29] M. Mundschau, E. Bauer, W. Swiech, *Catalysis Lett.* **1988**, *1*, 405.
[30] M. Mundschau, E. Bauer, W. Telieps, W. Swiech, *Surf. Sci.* **1989**, *213*, 38130.
[31] B. Rausenberger, W. Swiech, W. Engel, A. M. Bradshaw, E. Zeitler, *Surf. Sci.* **1993**, *287/288*, 235.
[32] W. Swiech, B. Rausenberger, W. Engel, A. M. Bradshaw, E. Zeitler, *Surf. Sci.* **1993**, *294*, 297.
[33] S. Jakubith, H. H. Rotermund, W. Engel, A. von Oertzen, G. Ertl, *Phys. Rev. Lett.* **1990**, *65*, 3013.
[34] M. S. Altman, E. Bauer, *Surf. Sci.* **1995**, *347*, 265.
[35] E. Bauer, Y. Wei, T. Müller, A. Pavlovska, I. S. T. Tsong, *Phys. Rev. B* **1995**, *51*, 17891.
[36] W. Swiech, E. Bauer, *Surf. Sci.* **1991**, *255*, 219.
[37] (a) M. Mundschau, E. Bauer, W. Swiech, *J. Appl. Phys.* **1989**, *65*, 581; (b) M. Mundschau, E. Bauer, W. Swiech, *J. Appl. Phys.* **1989**, *65*, 4747.
[38] I. Müllerova, M. Lenz, *Ultramicroscopy* **1992**, *41*, 399.
[39] I. Müllerova, L. Frank, *Scanning* **1993**, *15*, 193.
[40] L. Li, C. Koziol, K. Wurm, Y. Hong, E. Bauer, I. S. T. Tsong, *Phys. Rev.* **1994**, *50*, 10834.
[41] T. Michely, M. C. Reuter, M. Copel, R. M. Tromp, *Phys. Rev. Lett.* **1994**, *73*, 2095.
[42] R. M. Tromp, T. Michely, *Nature* **1995**, *373*, 499.
[43] M. S. Altmen, E. Bauer, *Surf. Sci.* **1995**, *344*, 51.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

1.7 Lorentz Microscopy

1.7.1 Magnetic Domains

Specimens of ferromagnetic materials are usually not magnetized uniformly throughout their volume, but are subdivided into domains. Within each domain, the magnetization is uniform, but different domains are magnetized in different directions. The domains are separated by boundaries called *domain walls*. Within a domain wall separating two domains, the magnetization direction gradually changes away from that in one domain towards that in the other.

The subdivision into domains is a result of the requirement for the free energy of the solid to be a minimum. The shape, size and arrangement of the domains, as well as the width of the domain walls, are determined by the four magnetic contributions to the total energy: the exchange, magnetostatic, anisotropy and magnetoelastic energies. When an external magnetic field is applied, the balance of the energy terms changes, resulting in a rearrangement of the domains. This rearrangement can take place by the displacement of domain walls (wall motion) or by rotation of the magnetization of the domains. Since these processes are to some extent irreversible, the change of average magnetization of the specimen as a function of the applied magnetic field is subject to hysteresis. The reader is referred to standard textbooks on magnetism for details. However, it is clear that the magnetic domain structure plays an important role in determining the magnetic properties, and it is therefore important to study the magnetic domain structure.

Many techniques have been developed for imaging magnetic domains. These techniques can be divided into the following categories:

(1) Techniques using small magnetic particles.
(2) Polarized light microscopy.
(3) Electron optical techniques.
(4) X-ray techniques.
(5) Neutron beam techniques.
(6) Scanning probe techniques.

None of the techniques is able to image the magnetic domains throughout the volume of a bulk specimen. Some techniques provide images of domains on the surfaces of specimens, and others can image domains in thin films. Among the electron optical techniques there are both types. In this article, we are concerned with the use of conventional transmission electron microscopes for domain imaging. The relevant techniques, which are collectively described as Lorentz microscopy, fall into the category of thin film techniques.

1.7.2 The Effect of a Magnetic Specimen on Electrons

Electrons carry an electric charge e, and when an electron passes through a region in which a magnetic induction $\boldsymbol{B}$ is present, it experiences a force. Classically, this force can be expressed as

$$\boldsymbol{F} = -|e|\boldsymbol{v} \times \boldsymbol{B} \tag{1}$$

provided no electric field is present, where $\boldsymbol{v}$ is the velocity of the electron. If an electron travels for a distance t through a region in which a magnetic induction $\boldsymbol{B}$ perpendicular to $\boldsymbol{v}$ is present, it will be deflected by an angle β given by

$$\beta = \frac{eBt}{mv} \tag{2}$$

where m is the mass of the electron. (Note that B and v are shorter forms of $|\boldsymbol{B}|$ and $|\boldsymbol{v}|$, respectively.) This equation assumes that β is small, which is always true in the cases we are concerned with. Another form of Eq. (2) is

$$\beta = \frac{eB\lambda t}{h} \tag{3}$$

where λ is the electron wavelength and h is Planck's constant. Equation (3) is in a convenient form even for very fast electrons, since relativistic effects can be taken into account by using the correct expression for λ.

An upper limit to β is set by the fact that B is limited by the properties of existing materials, and t is limited by the penetration of electrons. In practice, β is never more than about 10^{-4} rad. It is important to note that this value is about two orders of magnitude smaller than the Bragg diffraction angles from low-index atomic planes in typical metals. The presence of magnetic domains therefore only causes a small perturbation to the electron diffraction pattern due to the crystalline nature of specimens.

1.7.3 Methods of Observing Magnetic Domains in the TEM

Since the effect of the magnetic induction is to deflect the electrons, the intensity of an electron beam passing through a magnetic specimen is unaffected by the presence of magnetic domains. The presence of the magnetic induction only changes the direction in which the electrons emerge from the specimen, but not the number or even the speed of the electrons. A perfect, in-focus image of the specimen therefore contains no information about the domain structure. There are, however, three ways in which the presence of domains can be observed. The basic principles are all illustrated in Fig. 1. Figure 1a shows how the featureless in-focus image is formed. The top part of the figure shows the electron ray diagram, together with a cross-section of the specimen, the objective lens and the objective aperture. The specimen shown has domains magnetized perpendicular to the incident beam, alternately out of, and into, the plane of the paper (illustrated by the arrow heads and tails, respectively). The electrons are incident normally on the specimen, illuminating it uniformly. After they have passed through the specimen, the electrons are deflected either to the right or to the left, depending on the domain magnetization direction, and after passing through the objective lens they form an image projected onto the screen.

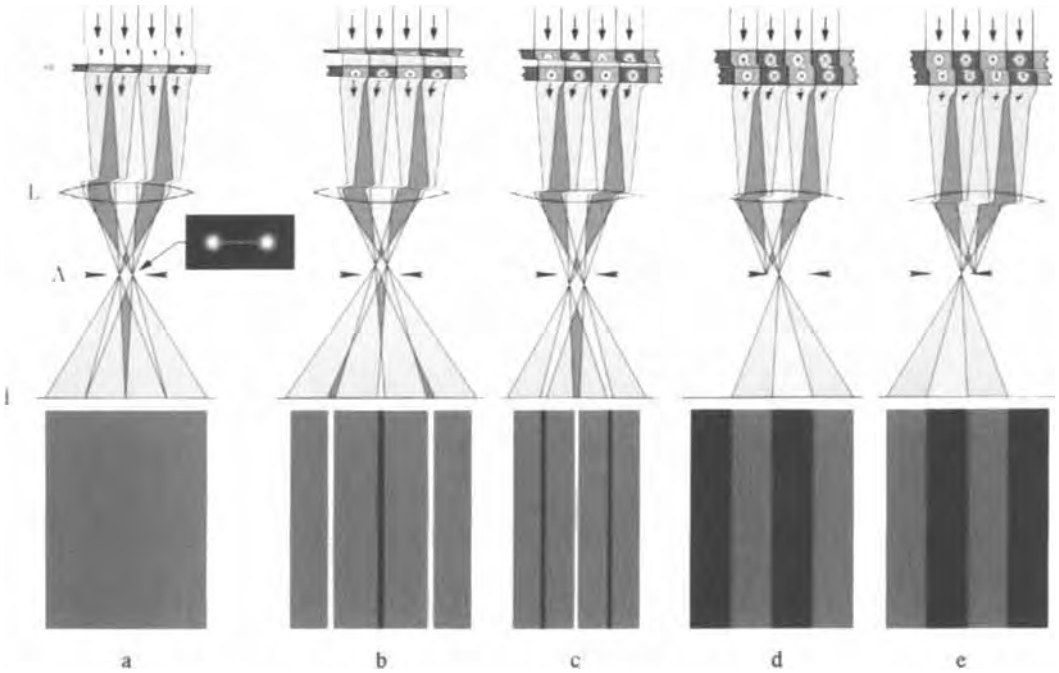


Figure 1. Schematic diagrams illustrating the principle of Lorentz microscopy. The upper part of each diagram shows the electron ray paths, and the lower part shows the type of image that would be obtained in different imaging conditions. The specimen S is illuminated by a parallel beam of electrons incident from above, travelling in the direction of the arrows. L is the objective lens, A the objective aperture, and I the image plane. In (a) the image of the specimen at I is in focus. The inset shows the appearance of the central diffraction spot. The objective lens is overfocused in (b) and underfocused in (c), and the objective aperture is displaced in opposite directions in (d) and (e).

However, the screen is uniformly illuminated, as illustrated schematically at the bottom of the figure. (Note that in a real electron microscope, there would be two or more additional lenses producing a highly magnified final image. There would not be a screen in the position indicated, but the image plane of the objective lens would be the object plane of the next lens below.)

Although the in-focus image shows no evidence of the presence of domains in the specimen, the domains have an effect on the diffraction pattern. Figure 1a includes a simulation of the central part of the diffraction pattern, in which the undiffracted spot (the spot formed by the part of the electron beam unaffected by the crystal structure of the specimen) has split into two subspots. These subspots correspond to the two magnetization directions present in the specimen.

In Fig. 1b the objective lens is overfocused, so that the image plane is slightly above the plane I. In the image shown in the bottom part of the figure, we now observe a dark line across the center, and two bright lines on either side. The ray diagram shows that these lines appear in

positions where the number of incident electrons has increased or decreased, respectively, as a result of the defocusing. The lines are in positions where the domain walls are imaged. The contrast reverses when the objective lens is underfocused, as shown in Fig. 1c. The image plane is now below the plane I, and in the plane I we observe a bright line in the center and two dark lines on either side. (Defocusing the objective lens alters the magnification of the image because of the change in the lens current, and not because of the magnetic effects of the specimen. These changes of magnification are somewhat exaggerated in Figs. 1b and 1c.)

A second method of imaging the domain structure is to displace the objective aperture from its position centered on the optic axis of the microscope, as shown in Figs. 1d and 1e. In Fig. 1d the aperture has been moved to the right, and it now blocks out electrons that have passed through domains magnetized into the paper (marked with crosses). The parts of the image corresponding to these domains now appear dark. In Fig. 1e the aperture has been moved to the left, blocking out electrons that have passed through domains magnetized out of the paper (marked with dots).

The defocus method of imaging domains is often called the *Fresnel method* (for reasons explained in Sec. 1.7.5 in this Chapter), and the displaced aperture method is usually called the *Foucault method*, since it is analogous to a method of testing the quality of optical components developed by Foucault in the 19th century. Use of the Fresnel method of Lorentz microscopy was first reported by Hale, Fuller and coworkers [1, 2], and the first reports of the use of the Foucault technique were published soon after [3, 4]. Reviews of Lorentz microscopy are given by Jakubovics [5] and Chapman [6].

1.7.4 Examples of Domain Images

Figure 2 shows transmission electron micrographs of a Co/Cu multilayer film, consisting of 40 cobalt layers of thickness 2 nm interleaved with copper layers of thickness 0.5 nm. Figure 2a is an in-focus image, which shows no evidence of a magnetic domain structure. The fine-scale contrast is due to the grain structure of the film. The inset shows the central part of the diffraction pattern, which is split into three main spots, indicating that there are three different magnetization directions in this region of the specimen. Figures 2b and 2c show images taken with the objective lens overfocused and underfocused, respectively. In Fig. 2b, three bright domain walls are visible, which become dark in Fig. 2c. The magnetization direction (shown by arrows in Fig. 2b) in the three domains can be determined on the assumption that the walls do not carry any magnetic charges (i.e., there is no change in the component of magnetization perpendicular to the wall across any of the walls). The magnetization directions are confirmed by the magnetization ripple (see Sec. 1.7.7). The presence of the ripple causes the three spots in the diffraction pattern to become slightly elongated. Figures 2d–2f show images obtained by displacing the objective aperture so that in each case only

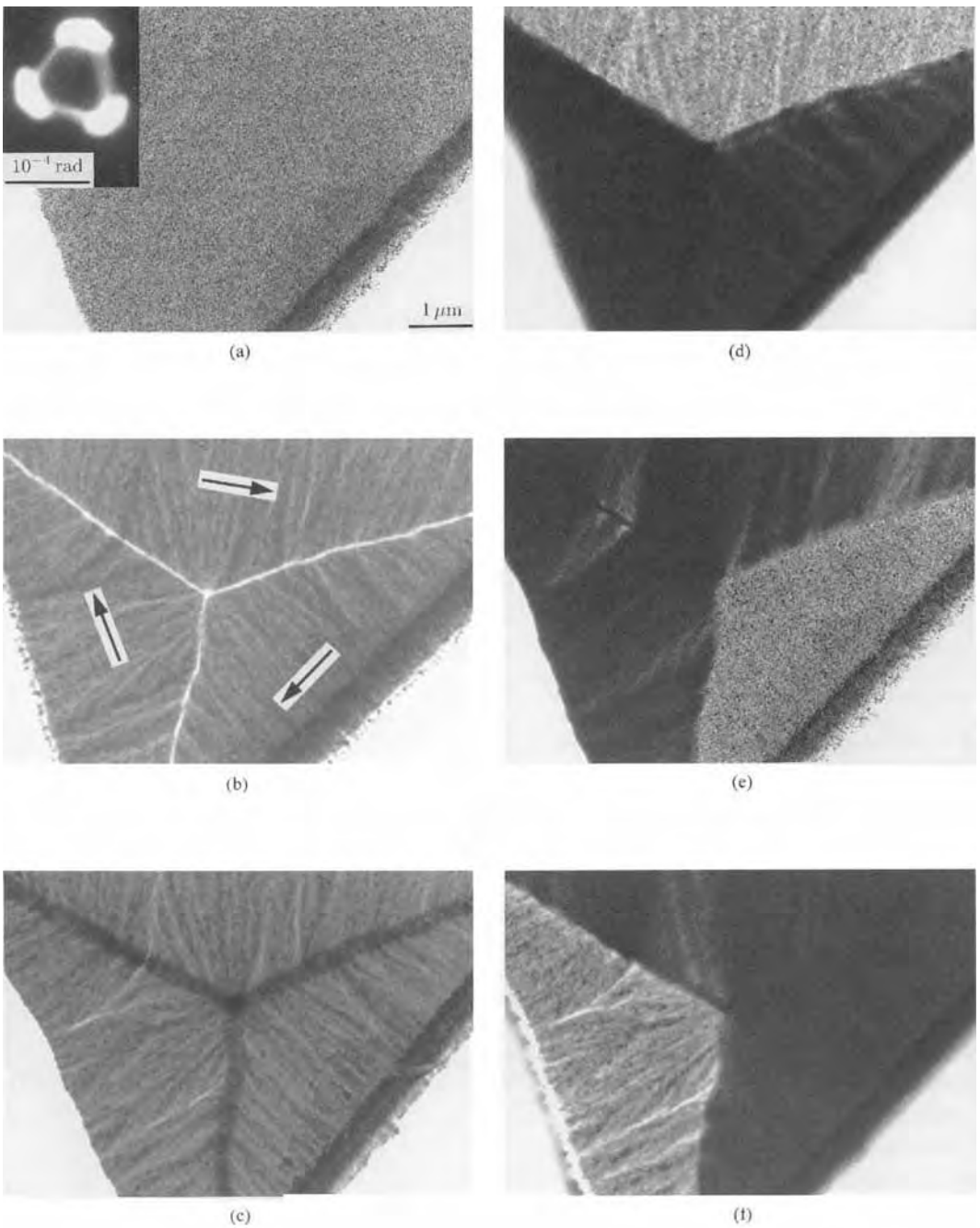


Figure 2. Micrographs of a $(Co_{2\,nm}/Cu_{0.5\,nm})_{40}$ multilayer film: (a) in focus; (b, c) with the objective lens overfocused and underfocused, respectively; (d–f) with the objective aperture displaced so that in each case, only one of the three spots in the central part of the diffraction pattern (shown as an inset in (a)) is inside the aperture. The micrographs were taken by J. D. Kim.

one of the three subspots seen in the inset in Fig. 2a is inside the aperture. In each of these Foucault images, one of the three domains is bright and the other two are dark.

1.7.5 Theory of Image Contrast

The first attempts to explain the contrast at domain walls in defocused images [2] were based on geometrical optics. It was also demonstrated that the wave nature of the electrons can have a strong effect on the image [7–9]: if the incident illumination is sufficiently coherent, a set of interference fringes appears in bright domain wall images. This effect is analogous to the interference fringes produced by a Fresnel biprism in light optics, and this is the reason why the defocus method of imaging domains in Lorentz microscopy is called the 'Fresnel method'. Subsequently, it was shown that the wave theory was generally applicable to Lorentz microscopy, and that the intensity distribution in both Fresnel and Foucault images could be calculated from the Abbe theory of microscopy [10]. It was also shown [11] that in certain circumstances, the geometrical theory could be used to obtain reasonable approximations to the results of the wave theory. These approximations have become less useful, because fast Fourier transform methods, which enable image intensities to be calculated rapidly from the wave theory, have become easily available. We therefore give a brief outline of the wave theory.

The deflection of electrons, Eq. (2), caused by the classical Lorentz force, is equivalent to a phase change. In the following discussion, we assume that the surfaces of the specimen lie in the xy plane, the incident electrons are travelling along the $-z$ direction, and the y component of the magnetic induction, B_y, is only a function of x. We denote the integral of B_y along the z direction by B_i. The phase difference between the electron wave at $x = 0$ and $x = X$ is

$$\phi(X) = -\frac{|e|t}{\hbar}\int_0^X B_i(x)\,dx \tag{4}$$

where $\hbar = h/2\pi$. If the magnetic induction is confined to the volume of the specimen, t can be taken to be the specimen thickness. According to the Abbe theory, the amplitude in the back focal plane of the objective lens (diffraction plane) is the Fourier transform of the transmission function of the specimen, and it can be expressed as

$$A_d(\theta) = \int_{-\infty}^{\infty} \times \exp\left\{i\left[\phi(x) - \frac{2\pi(\theta+\theta_i)x}{\lambda}\right]\right\} dx \tag{5}$$

where θ is the angular variable in the diffraction plane parallel to x, assuming that the specimen is illuminated by a plane wave of electrons arriving at an angle θ_i to the normal to the specimen surface. The amplitude of the diffraction pattern is, in general, modulated by a complex function $f(\theta)$, which includes the effects of defocus, finite objective aperture size and lens aberrations. The amplitude $A_i(\xi)$ in the image is the Fourier transform

of $A_d(\theta)f(\theta)$:

$$A_i(\xi) = \int_{-\infty}^{\infty} A_d(\theta)f(\theta) \times \exp\left(-\frac{2i\pi\theta\xi}{\lambda}\right) d\theta \qquad (6)$$

where ξ is a coordinate in the image plane parallel to x. (Note that, since the limits of the integral are infinite, we should really use $\tan\theta$ in the integrand. However, as any appreciable contribution to the amplitude comes from small θ, we can neglect the difference between θ and $\tan\theta$.) The image intensity is then given by

$$I_i(\xi) = |A_i(\xi)|^2 \qquad (7)$$

The effect of a finite source size can be taken into account by performing a convolution with the intensity of illumination arriving in different directions.

The imaging modes in Lorentz microscopy can be described by using the general function

$$f(\theta) = a(\theta)\exp\left[-\frac{i\pi}{\lambda}(\tfrac{1}{2}C_s\theta^4 + \zeta\theta^2)\right] \qquad (8)$$

where $a(\theta)$ is the aperture function, C_s is the spherical aberration coefficient and ζ is the defocus, with $\zeta > 0$ corresponding to overfocus. In the one-dimensional case so far considered, we take $a(\theta) = 1$ for $\theta_1 \leqslant \theta \leqslant \theta_2$ and $a(\theta) = 0$ elsewhere, where θ_1 and θ_2 correspond to the edges of the aperture. The two special cases of interest are: (i) the Fresnel case, in which we neglect the aperture, that is we take $a(\theta) = 1$ for all θ; and (ii) the Foucault case, in which $\zeta = 0$ and either θ_1 or θ_2 is finite. The effects of spherical aberration are usually negligible, because of the small values of the Lorentz deflection.

The above equations can easily be extended to two dimensions, that is to the case where the magnetic induction varies in both the x and y directions [6].

1.7.6 Interference Effects

Figure 3 shows an area containing a cross-tie wall [12, 13] in a polycrystalline cobalt film of thickness 67 nm. Figures 3a and 3b show two images of the same area taken with the objective lens overfocused and underfocused, respectively. Figures 3c and 3d show the part of the area circled in Fig. 3a, taken with a more parallel incident beam, that is with a more coherent illumination. It is seen that the bright domain walls in each image now break up into interference fringes.

1.7.7 Determination of Magnetization Distributions

In general, there is no simple relationship between the intensity distribution in Fresnel or Foucault images and the magnetization distribution or even the magnetic induction distribution in the specimen. In a specimen containing uniformly magnetized domains separated by domain walls that are narrow compared with the size of the domains, the Fresnel method can be used to find the position of the domain walls, and the Foucault method can be used to find the direction of magnetization of the domains. In cases where the domains are not uniformly magnetized but contain magnetization ripple [2], the

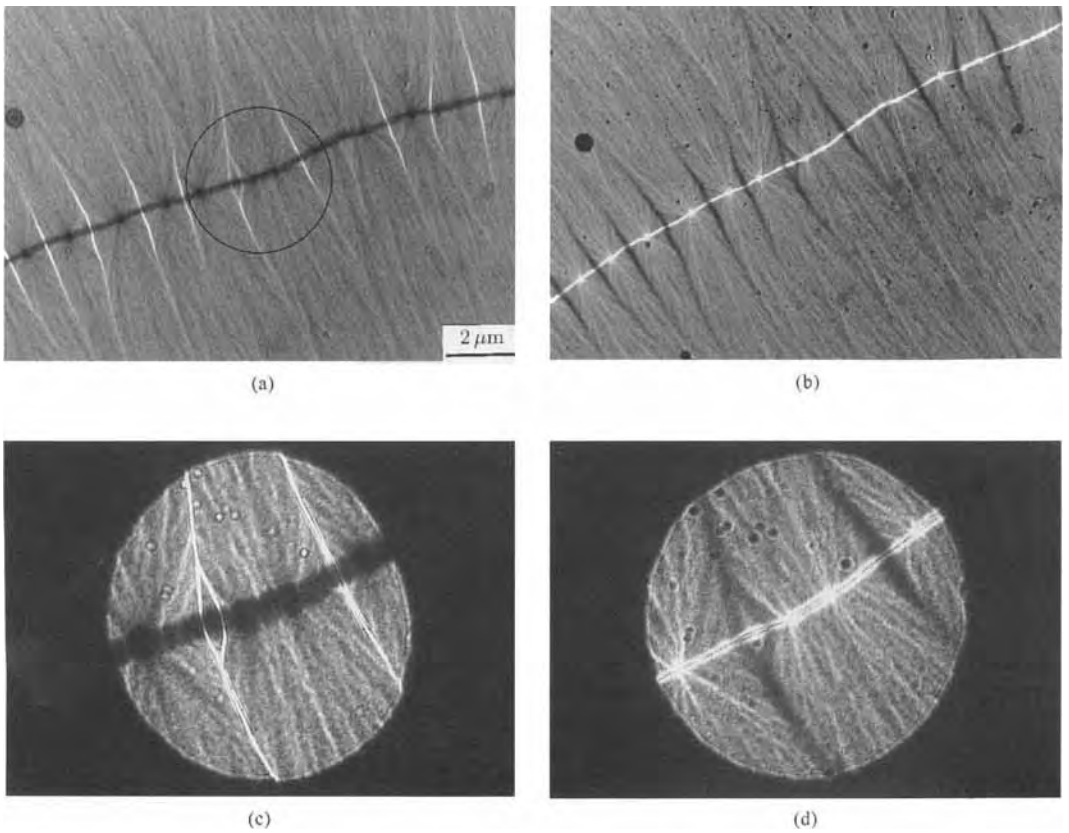


Figure 3. Micrographs of a polycrystalline cobalt film of thickness 67 nm. The objective lens is overfocused in (a) and underfocused in (b). Micrographs of the area circled in (a), taken with highly coherent illumination and with the objective lens defocused in opposite directions, are shown in (c) and (d). The micrographs were taken by J. D. Kim.

magnetization direction of the domains can be determined from either Fresnel or Foucault images, since the ripple is perpendicular to the magnetization (Fig. 2). However, details of the magnetization distribution, such as the structure of domain walls, cannot be deduced directly from Lorentz micrographs. At best, it is possible to determine domain wall widths by matching the observed intensity distribution to a series of computed profiles (e.g., [14]), or to detect the well-known asymmetry [15, 16] of some domain walls in thin films [17–19]. Magnetization distributions can be mapped quantitatively by using a suitably modified scanning transmission electron microscope (TEM) [20], but this topic is outside the scope of the present article. It has, however, been shown recently [21] that an equivalent mapping method can be implemented in a conventional TEM by combining series of Foucault images. Such methods are becoming feasible with the use of digital image-acquisition facilities and computer control of electron microscopes. Other new methods for mapping magnetization distributions by transmission electron

microscopy have also been developed recently [22].

1.7.8 Practical Implementation of Lorentz Microscopy

Magnetic lenses used in TEMs produce magnetic fields approximately parallel to the axis of the microscope, which are strong enough to saturate most specimens. In order to study magnetic domains, it is therefore necessary to reduce the magnetic field acting on the specimen. In some cases, acceptable results can be achieved by switching off the objective lens and using the remaining lenses to produce an image, but the magnification and resolution available are usually insufficient. In TEMs with top-entry specimen stages, it is sometimes possible to achieve a sufficient reduction of the magnetic field by moving the specimen away from the objective lens and increasing the focal length of the lens. In TEMs with side-entry specimen stages, it is not possible to alter the specimen position. However, specially designed objective pole-pieces are available for some TEMs, which concentrate the focusing fields well below the specimen, and thus enable any standard side-entry specimen holder to be used [22, 23]. These special pole-pieces can achieve high magnification and resolution.

Reduction of the magnetic field acting on the specimen is the only modification to a standard TEM needed for Fresnel imaging. For Foucault imaging, there is the additional difficulty that if the focal length of the objective lens is increased, the objective aperture is no longer in the back-focal plane. When the objective aperture is displaced, its edge casts a fairly narrow shadow over the screen, and the conditions to obtain Foucault contrast are only satisfied within this shadow. To overcome this problem, it is necessary to move the plane of the aperture away from the lens. Doole et al. [24] give an example of a successful solution to this problem.

It is sometimes useful to combine Lorentz imaging with in situ experiments. The use of side-entry specimen holders makes such experiments particularly convenient, since this arrangement enables all standard specimen holders, such as heating and cooling holders, to be used. The study of magnetic specimens often requires the application of controlled magnetic fields, in order to study the motion of domain walls. However, applied magnetic fields affect not only the specimen, but also the electron beam. A field applied in the plane of the specimen deflects the beam, since the field acts as an electron optical prism. Although it is possible to compensate for this deflection by applying magnetic fields in the opposite direction away from the specimen, the fields also introduce aberrations, which ultimately limit the magnitude of the field that can be applied. For a successful design of a set of in situ magnetizing coils for use with a top-entry magnetizing stage, see Taylor [25]. If it is required to apply magnetic fields perpendicular to the specimen, the field of the objective lens can sometimes be used, the specimen being moved along the axis of the microscope [26].

Acknowledgements

The author is grateful to J. D. Kim for the micrographs shown in Figs. 2 and 3, and to A. K. Petford-Long for helpful comments on the manuscript.

1.7.9 References

[1] M. E. Hale, H. W. Fuller, H. Rubinstein, *J. Appl. Phys.* **1959,** *30*, 789.

[2] H. W. Fuller, M. E. Hale, *J. Appl. Phys.* **1960,** *31*, 238.

[3] H. Boersch, H. Raith, *Naturwissenschaften* **1960,** *46*, 576.

[4] H. W. Fuller, M. E. Hale, *J. Appl. Phys.* **1960,** *31*, 1699.

[5] J. P. Jakubovics, in *Electron Microscopy in Materials Science, Third Course of the International School of Electron Microscopy* (Eds.: E. Ruedl, U. Valdrè), Commission of the European Communities **1975,** 1303.

[6] J. N. Chapman, *J. Phys. D: Appl. Phys.* **1984,** *17*, 623.

[7] H. Boersch, H. Hämisch, D. Wohlleben, K. Grohmann, *Z. Phys.* **1960,** *159*, 397.

[8] H. Boersch, H. Hämisch, D. Wohlleben, K. Grohmann, *Z. Phys.* **1961,** *164*, 55.

[9] H. Boersch, H. Hämisch, K. Grohmann, D. Wohlleben, *Z. Phys.* **1960,** *167*, 72.

[10] D. Wohlleben, *J. Appl. Phys.* **1967,** *38*, 3341.

[11] J. P. Guigay, R. H. Wade, *Phys. Stat. Sol.* **1968,** *29*, 799.

[12] E. E. Huber, D. O. Smith, J. B. Goodenough, *J. Appl. Phys.* **1958,** *29*, 294.

[13] S. Methfessel, S. Middelhoek, H. Thomas, *IBM J. Res. Dev.* **1960,** *4*, 96.

[14] D. C. Hothersall, *Phil. Mag.* **1969,** *20*, 89.

[15] A. E. LaBonte, *J. Appl. Phys.* **1969,** *40*, 2450.

[16] A. Hubert, *Phys. Stat. Sol.* **1969,** *321*, 519.

[17] D. C. Hothersall, *Phys. Stat. Sol. B* **1972,** *51*, 529.

[18] C. G. Harrison, K. D. Leaver, *Phys. Stat. Sol. A* **1972,** *12*, 413.

[19] C. G. Harrison, K. D. Leaver, *Phys. Stat. Sol. A* **1973,** *15*, 415.

[20] J. N. Chapman, R. Ploessl, D. M. Donnet, *Ultramicroscopy* **1992,** *47*, 331.

[21] A. C. Daykin, A. K. Petford-Long, *Ultramicroscopy* **1995,** *58*, 365.

[22] J. N. Chapman, A. B. Johnston, L. J. Heyderman, S. McVitie, W. A. P. Nicholson, B. Bormans, *IEEE Trans. Magn.* **1994,** *30*, 4479.

[23] K. Tsuno, M. Inoue, K. Ueno, *Mat. Sci. Eng. B* **1989,** *3*, 403.

[24] R. C. Doole, A. K. Petford-Long, J. P. Jakubovics, *Rev. Sci. Instrum.* **1993,** *64*, 1038.

[25] R. A. Taylor, in *Electron Microscopy* **1980,** Vol. 4, Proc. 6th Int. Conf. on High Voltage Electron Microscopy, Antwerp (Eds.: P. Brederoo, J. Van Landuyt), Seventh European Congress on Electron Microscopy Foundation, Leiden **1980,** 38.

[26] E. L. Houseman, J. P. Jakubovics, *J. Magn. Magn. Mater.* **1983,** *31*, 1007.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

1.8 Electron Holography Methods

1.8.1 Principle and Problems of Conventional Transmission Electron Microscopy

1.8.1.1 Modulation of the Electron Wave

As early as 1932, Glaser [1] showed that, for electrons moving with an energy eU in a medium with an electrostatic potential field $V_{el}(\boldsymbol{r})$ and a magnetic potential field $\boldsymbol{A}(\boldsymbol{r})$, the index of refraction is

$$n\left(\boldsymbol{r}, \frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}s}\right) = \sqrt{\frac{[U + V_{el}(\boldsymbol{r})]^*}{U^*}} - \sqrt{\frac{e}{2m_0 U}}\,\boldsymbol{A}(\boldsymbol{r})\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}s} \tag{1}$$

Evidently, the electric contribution depends only on the position $\boldsymbol{r}$, whereas the magnetic potential contribution depends on both $\boldsymbol{r}$ and the direction $\mathrm{d}\boldsymbol{r}/\mathrm{d}s$ of the electron path. Comparing two paths s_1 and s_2 through a medium, an electron wave undergoes a phase shift

$$\phi = \frac{2\pi}{\lambda_0}\left\{\int_{s_2} n\,\mathrm{d}s - \int_{s_1} n\,\mathrm{d}s\right\} \tag{2}$$

Electric Potential, $V_{el}(\boldsymbol{r})$

In the case of the electric potential distribution $V_{el}(\boldsymbol{r})$ representing an object, the phase shift with respect to vacuum is given by

$$\phi = \frac{2\pi}{\lambda_0}\int_{\mathrm{obj}} (n-1)\,\mathrm{d}s \tag{3}$$

where the integral has to be taken along the electron path s.

Since $V \ll U$, this can be written as

$$\phi = \frac{2\pi}{\lambda_0}\frac{1}{2U^*}\frac{m}{m_0}\int_{\mathrm{obj}} V(\boldsymbol{r})\,\mathrm{d}z \tag{4}$$

or

$$\phi = \sigma V_{\mathrm{p}} \tag{5}$$

with the interaction parameter

$$\sigma = \frac{2\pi\lambda_0 em}{h^2} \tag{6}$$

and the projected potential

$$V_{\mathrm{p}}(x,y) = \int_{\mathrm{obj}} V_{el}(\boldsymbol{r})\,\mathrm{d}z \tag{7}$$

The projected potential may represent a single atom, an amorphous thin film, or a crystal that is thin compared to the extinction distance. In thicker crystals, the effects of dynamic interaction have to be taken into account, which need, for example, multislice calculations instead of the simple potential projection.

Magnetic Potential, $\boldsymbol{A}(\boldsymbol{r})$

Integration of

$$\phi = \frac{2\pi}{\lambda_0}\sqrt{\frac{e}{2m_0U^*}} \times \left[\int_{s_2} \boldsymbol{A}(\boldsymbol{r})\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}s}\,\mathrm{d}s - \int_{s_1} \boldsymbol{A}(\boldsymbol{r})\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}s}\,\mathrm{d}s\right] \quad (8)$$

with $\boldsymbol{B} = \mathrm{rot}\,\boldsymbol{A}$ and the magnetic flux $\Phi_m = \oint \boldsymbol{A}\,\mathrm{d}\boldsymbol{r}$ enclosed in the closed path s yields

$$\phi = \frac{e}{\hbar}\Phi_m \quad (9)$$

This phase shift occurs even if the electron path does not lead through an area with a magnetic field $\boldsymbol{B}$, that is even if no Lorentz force is acting on the electrons. This *Aharonov–Bohm effect* was discovered theoretically by Ehrenberg and Siday [2] and by Aharonov and Bohm [3].

An electron wave transmitted through an object undergoes a phase shift $\phi(x,y)$ according to the electric or magnetic potential distribution. In addition, the amplitude $a(x,y)$ may be modulated due to the following causes: scattering into angles larger than the collection angle of the subsequent optics, interference effects at the exit face of the object and to inelastic scattering (inelastically scattered electrons are no longer coherent with the elastically scattered ones).

In conclusion, the electron wave at the exit face of an object is given by the 'object function'

$$\mathrm{o}(x,y) = a(x,y)\cdot\exp[\mathrm{i}\phi(x,y)] \quad (10)$$

1.8.1.2 Propagation of the Electron Wave through the Electron Microscope

The electron wave leaving the object propagates in space according to the wave equation, thereby changing its appearance in amplitude and phase in a rather complicated manner. Nevertheless, once the wave is known completely over a sufficiently wide surface it can be determined at every point in space by means of the Kirchhoff diffraction integral (KDI). Most familiar special cases frequently used in optics are the Fresnel approximation of the KDI delivering the Fresnel diffraction pattern, or, more importantly, the far field approximation of the KDI which allows one to determine the wave at the Fraunhofer diffraction condition by means of a Fourier transform.

The object wave propagates through the electron microscope according to the wave equation. Therefore, taking into account the properties of the imaging system, one can track the wave from the object to the image plane. In general, the image wave may be written as

$$b(x,y) = \int \mathrm{o}(\tilde{x},\tilde{y})G(\tilde{x},\tilde{y},x,y)\,\mathrm{d}\tilde{x}\,\mathrm{d}\tilde{y} \quad (11)$$

Usually, the imaging process is considered isoplanar, that is $G(\tilde{x},\tilde{y},x,y)$ depends only on the differences $(\tilde{x}-x)$ and $(\tilde{y}-y)$, and hence the image wave is found as a convolution of the object wave with the 'point spread function' (PSF) $(x,y) = G(0,0,x,y)$

$$b(x,y) = \mathrm{o}(x,y)\otimes\mathrm{PSF}(x,y) \quad (12)$$

In Fourier space this reads

$$\mathrm{FT}(b) = \mathrm{FT}(\mathrm{o})\cdot\mathrm{WTF}(\boldsymbol{R}) \quad (13)$$

$\mathrm{WTF}(\boldsymbol{R}) = \mathrm{FT}^{-1}(\mathrm{PSF})$ is the wave transfer function, which depends on the spatial frequency vector $\boldsymbol{R} = (R_x, R_y)$. With an aberration-free imaging system, $\mathrm{WTF}(\boldsymbol{R}) \equiv 1$ holds, and hence the PSF would be a delta function (i.e., an

arbitrarily sharp point). Then, apart from a magnification factor, the image wave and object wave would match perfectly in both amplitude and phase.

However, electron lenses cannot be made free from aberrations; most severe are the spherical aberrations, which are unavoidable with the usual electron lenses [4], and chromatic aberration. In particular at increasingly high resolution there are a variety of additional aberrations to be considered, which can be classed as either coherent or incoherent aberrations.

Coherent aberrations are those aberrations which may be present under perfectly coherent illumination. These are spherical aberrations, defocus, astigmatisms and axial coma. They enter the imaging process by means of a wave aberration $\chi(\boldsymbol{R})$ which describes the resulting deviation of the wavefront in Fourier space from the ideal one. Coherent aberrations act like a phase plate in Fourier space. They do not destroy information. To understand their effect, one must distinguish between aberrations which are symmetrical and those which are antisymmetrical in $\boldsymbol{R}$. Those *symmetrical* in $\boldsymbol{R}$ (i.e., $\chi_s(\boldsymbol{R}) = \chi_s(-\boldsymbol{R})$), such as spherical aberrations, defocus and two-fold astigmatism, act like a Zernike phase plate with a thickness which is variable in $\boldsymbol{R}$. Consequently, symmetric aberrations exchange amplitudes and phases in the image wave with respect to the object wave according to the $\sin[\chi_s(\boldsymbol{R})]$ and $\cos[\chi_s(\boldsymbol{R})]$ functions. The aberrations *antisymmetric* in $\boldsymbol{R}$ (i.e., $\chi_a(\boldsymbol{R}) = -\chi_a(-\boldsymbol{R})$), such as three-fold astigmatism and axial coma, produce a lateral shift $\chi_a(\boldsymbol{R})$ of the elementary cosinoidal gratings associated in real space with each spatial frequency; this may severely degrade the image wave.

Incoherent aberrations always show up due to the restricted coherence of electron illumination. Due to the illumination aperture θ_c (i.e., lateral coherence) and the energy spread ΔE (i.e., longitudinal coherence) the envelope functions $E_{coh}(\boldsymbol{R})$ and $E_{chrom}(\boldsymbol{R})$ dampen the Fourier spectrum of the image wave. Incoherent aberrations act like an aperture in Fourier space cutting away the contributions of the higher spatial frequencies, and hence destroy information in that they prevent object information from showing up in the image plane.

In summary, the wave transfer function is given by

$$\mathrm{WTF}(\boldsymbol{R}) = E_{coh}(\boldsymbol{R}) \cdot E_{chrom}(\boldsymbol{R}) \cdot \exp[-i\chi(\boldsymbol{R})] \tag{14}$$

with

$$\chi(\boldsymbol{R}) = \chi_s(\boldsymbol{R}) + \chi_a(\boldsymbol{R}) \tag{15}$$

1.8.1.3 Problems in Conventional Transmission Electron Microscopy

A conventional electron image represents the intensity distribution given by

$$I(x,y) = b(x,y) \cdot b^*(x,y) \tag{16}$$

which is falsified with respect to the image wave $b(x,y)$ in several respects.

$I(x,y)$ only represents the squared amplitude of the image wave, whereas the phase of the image wave is lost in a micrograph. Therefore the desirable object information should be directed into the amplitude of the image by means of a suitable selection of the wave aberration. Generally, however, the information found in the image amplitude is a mixture

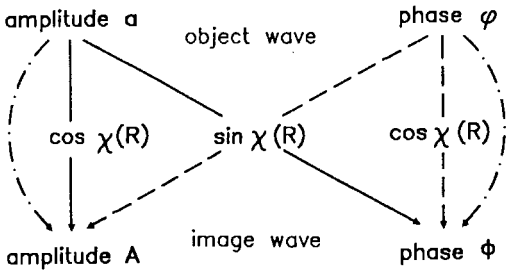


Figure 1. Schematic transfer of the object wave into the image wave by the objective lens with a wave aberration $\chi(R)$ symmetric in R. The Scherzer focus optimizes the $\sin[\chi(R)]$ paths. Aberration-free imaging means $\cos\chi \equiv 1$ at vanishing 'cross-talk' $\sin\chi \equiv 0$.

of both object amplitude and object phase transferred according to the transfer functions $\cos[\chi_s(\boldsymbol{R})]$ and $\sin[\chi_s(\boldsymbol{R})]$, respectively (Fig. 1). Therefore the image intensity cannot be interpreted uniquely in terms of object amplitude or phase. Only if a weak phase object is assumed does $I(x, y)$ represent solely the object phase—but distorted by the *phase contrast transfer function* $\sin[\chi_s(\boldsymbol{R})]$ (Fig. 2). Under

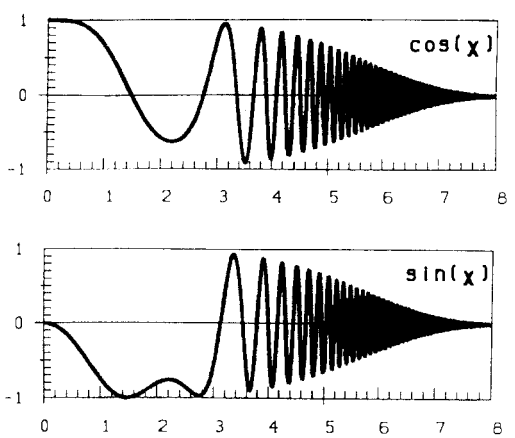


Figure 2. Wave transfer function $\mathrm{WTF}(\boldsymbol{R}) = E_{coh}(\boldsymbol{R}) \cdot E_{chrom}(\boldsymbol{R}) \cdot \exp[-i\chi(\boldsymbol{R})]$ at Scherzer focus. The point resolution is defined by the first zero of the $\sin\chi$ function. The information limit at about $6\,\mathrm{nm}^{-1}$ corresponds to a damping to $\exp(-1)$. (100 kV, energy width 0.5 eV, $C_c = 1.2$ mm, $C_s = 1.2$ mm, illumination aperture 0.05 mrad.)

the assumption that only the spherical aberration (coefficient C_s) and defocus Dz contribute to the wave aberration, the distortion can be minimized by selecting the Scherzer focus; then the function $\sin[\chi_s(\boldsymbol{R})]$ has a shape such that it approaches unity over a broad range of spatial frequencies ranging from $R_{min} = 0.385/(C_s^{0.25}\lambda_0^{0.75})$ to $R_{scherz} = 1.5/(C_s^{0.25}\lambda_0^{0.75})$ [5]. Therefore, large-area phase structures with $R < R_{min}$ are not visible. Spatial frequency components with $R > R_{scherz}$ cannot be interpreted intuitively because of the contrast reversals and the zeros; therefore R_{scherz} is called the *Scherzer point-resolution limit* of the electron microscope.

Nonlinearity also degrades the image information. Whereas, even under aberrated imaging conditions, the image wave is always linearly related to the object wave, this is not true in general for the image intensity. Due to the squaring of the amplitude, the intensity contains spatial frequencies which are not present in the object wave; the most common example of such frequencies are the half-spacings observed with strongly scattering crystals. They arise due to mutual interference of the reflections in addition to the regular interference of each reflection with the undiffracted beam. In addition, the inelastically scattered electrons contribute a background to the high-resolution image which is not included in the above description. The interpretation of such micrographs is severely impeded.

As the use of modern field emission guns greatly reduces the effect of the incoherent aberrations, an information limit of $R_{lim} \approx 2R_{scherz}$ is found. Of course, the point resolution R_{scherz} should now be expanded to the information limit R_{lim} by

reducing C_s, which is difficult to achieve. Besides, the problem of large-area phase contrast is worsened because the Scherzer band is shifted to higher spatial frequencies. In a fully corrected imaging system, no phase contrast shows up at all. Therefore, aberration-free imaging is only desirable in combination with well-elaborated wave optical techniques for analysis of the image wave to extract all accessible information about the object structure. However, the wave optical facilities in the electron microscope are relatively poor in that the parameter accessible to the operator is defocus. It is this state of affairs that holography improves considerably.

1.8.2 Holography: Basic Principles

The above described problems in conventional electron microscopy arise mainly because the wave optical imaging process is irreversibly brought to an end by recording the image intensity and thus intrinsically losing the image phase. Since the image phase information is missing, an a posteriori correction (e.g., of focus or astigmatism) in a micrograph is absolutely impossible. In order to accomplish this, we have to consider the wave optics: if the complete image wave was available in terms of amplitude and phase then, again according to the laws of wave optics, it could be back-propagated through an optical system to recover the object wave again.

Gabor [6] extended this idea to the scheme of holography, showing how a wave could be recorded and then 're-created' from the record: if one superimposes a coherent reference wave r on the image wave b, the resulting interference pattern represents a hologram given by the intensity distribution

$$I_{\text{hol}} = (r + b)(r + b)^* \tag{17}$$

Considering the hologram recorded on a photographic plate as a light optical diffraction grating, Gabor found one of the diffracted waves to be a copy wave of the recorded one:

> Thus, if one illuminates the photographic record with an optical imitation of the electronic (reference) wave, only that part of the primary (image) wave will be strongly transmitted which imitates the modified wave both in phases and amplitudes. (Gabor, 1948 [6])

In fact, under illumination with the reference wave r, three waves leave the hologram:

$$r \cdot I_{\text{hol}} = r(|r|^2 + |b|^2) + r^2 b^* + |r|^2 b \tag{18}$$

where the last term is a copy of the recorded wave b.

Once the wave b is re-created in the plane where the hologram was taken, it can be propagated to any desirable plane (e.g., back to the level of the object). The point is that the recorded wave can be propagated by means of any medium obeying the wave equation; for example, the electron wave can be re-created and propagated as a light wave which, apart from a different wavelength, obeys the same wave equation. Thereby one can combine the advantages of electron microscopy (i.e., short electron wavelength allowing high resolution and strong interaction of the electrons at atomic

dimensions) with those offered by well-established light optics for highly sophisticated wave optical handling (e.g., with the Zernike phase contrast method) and an analysis of amplitude and phase.

During the subsequent years of development of light-optical holography it was shown that the wave optical analysis of the object wave can be performed on the reconstructed wave (e.g., phase contrast [7], holographic interferometry [8], phase enhancement [9], and a posteriori focusing and correction of aberrations by means of a corresponding lens [10]).

1.8.3 Electron Holography: Aspects of Realization

The general scheme for electron holography is outlined in Fig. 3. The electron wave found on the final screen of an electron microscope is recorded by a detector and transferred to the reconstruction device. There the wave is re-created and back-propagated through an optical system which is equivalent to the electron microscope. Finally, the wave can be analyzed at the stage of the object and of its Fourier spectrum.

The key is to take high quality holograms in the first instance. This allows one to re-create the captured wave to a high degree of fidelity for further processing. Therefore, the electron optical facilities for doing wave optics with electrons had to, and still have to, be developed and optimized for this purpose, and restrictions peculiar to electron optical technology have to be considered.

1.8.3.1 Coherence

In light optics, holographic methods did not reach full success until the laser, as a perfectly coherent light source, became

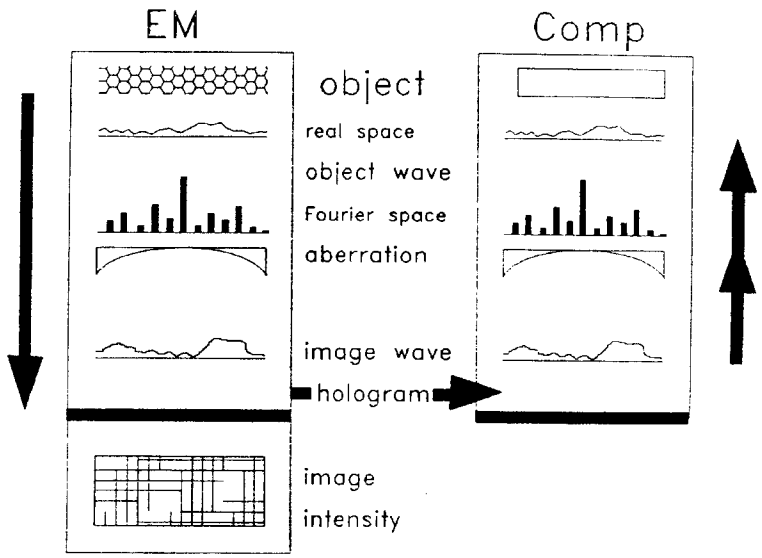


Figure 3. Scheme of electron holography. In the electron microscope, instead of the intensity, the complex electron wave is recorded in a hologram and transferred to the reconstruction device, which may consist of a light optical bench or a computer. There the wave is re-created and back-propagated to the Fourier plane or the object plane for further wave optical analysis.

available, because a well-defined phase relation between the wavefield to be recorded and the reference wave is indispensable. Indeed, for electron holography, the key point is the coherence of electron radiation which, as electron sources are incoherent, can only be achieved by means of the usual coherence conditions.

Because of the chromatic aberration of electron lenses, temporal coherence is essential because it dampens the wave transfer function of the electron microscope. Fortunately, from the point of view of interferometry and holography, this poses virtually no additional restriction since the coherence length amounts to several 10000 wavelengths [11]. Path-length differences in electron interference experiments, however, seldom exceed about 1000 wavelengths.

The main problem occurs with lateral coherence. Using incoherent sources, a plane at a large distance may be illuminated coherently so that, if two previously separate points P_1 and P_2 are superimposed, an interference contrast $0 < V < 1$ is found; usually, the contrast decreases with increasing distance between P_1 and P_2. V is the modulus of the degree of coherence which, according to the van Cittert–Zernike theorem, is distributed in the illuminated plane according to the Fourier transform of the intensity distribution of the source. In combination with the brightness B_0 of the electron gun one finds that the total coherent current available for producing an interference pattern with contrast V is limited by [12]

$$I_{coh} = -B_0 \cdot \lambda_0^2 \ln(V) \qquad (19)$$

As a property of the electron gun, I_{coh} is optically invariant along the microscope column.

Taking holograms at a sufficient contrast V, the current density is accordingly small. Therefore, to collect the electron dose needed for a low noise level, the exposure time is considerably higher than in conventional electron microscopy; this, in turn, is limited by residual instabilities, again reducing V. In any case, high brightness electron guns (e.g. field emission guns which have a brightness about 1000 times higher than thermionic guns) are indispensable.

1.8.3.2 Geometries for taking Electron Holograms

In the general formula describing a hologram

$$I_{hol} = (r + b)(r + b)^* \qquad (20)$$

it is left open in which way the object-modulated wave b and the reference wave r are superimposed, and where the hologram is taken.

The first point to consider is whether the two waves are running essentially parallel (in-line holograms) or are superimposed at some angle (off-axis holograms). This determines how the reconstructed waves leaving the hologram

$$r \cdot I_{hol} = r(|r|^2 + |b|^2) + r^2 b^* + |r|^2 b \qquad (21)$$

are mutually distributed in space. If r and b have been superimposed in parallel, then $r^2 = |r|^2$ holds and the three reconstructed waves travel essentially in the same direction. Here one faces the famous 'twin-image problem': b and b^* overlap and cannot be separated, and hence disturb each other unless the field of view is very

small. Furthermore, certain restrictions apply in that the interaction with the object should be sufficiently small to ensure that the (undiffracted) reference wave, which is also partly transmitted through the object, is not modulated.

For these reasons, holography was brought a huge step forward by Leith and Upatnieks [13], who proposed off-axis holography, that is where r and b are superimposed at some angle. The reconstructed b and b^* waves are then separated from the remainder by the overlapping angle in opposite directions. One wave, b or b^*, can be cut out and back-propagated through the optical system.

The next point to consider is where the holograms should be recorded. In principle, the hologram can be recorded anywhere in the wavefield, for example in the image plane (image plane holograms), at some distance (Fresnel holograms), at a large distance (Fraunhofer holograms) or in the back focal plane of a lens (Fourier holograms). In any case, after corresponding back-propagation, the object wave should look the same, irrespective of where the hologram was taken. The main criterion for taking the hologram is given by the achievable resolution. The PSF containing all the information about one object point spreads into large areas, the area increasing with the distance to the object. If the PSF exceeds the coherent domain of the reference wave it is no longer recorded completely in the hologram, and hence the wave cannot be reconstructed at full resolution.

Reviewing the holographic arrangements realized in electron optics, in 1992 Cowley listed 20 forms of electron holography [14]. Some of these are taken into account below.

1.8.3.3 In-Line Holography

Using the above described terminology, the first electron holography experiments (by Haine and Mulvey in 1952) were conducted as in-line Fresnel holography [15]. Their pioneering work verified the basic idea of electron holography; however, it suffered from the comparably poor electron optics technology, and the results revealed the problems of this technique (see above). Meanwhile, using a very coherent nanotip field emission source, Fink and coworkers [16, 17] have taken up this technique again, hoping to get rid of the twin-image by sophisticated numerical reconstruction; a detailed analysis of this method is given by Spence et al. [18].

In-line Fraunhofer holography was realized by Tonomura et al. [19], showing the value of this technique for holographic reconstruction in that the twin-image problem is avoided. More recent applications [20–22] show a resolution approaching 1 nm. However, this technique is restricted to weak and very small objects (e.g., to certain biological objects) [23].

Fourier holography has been realized experimentally [24], but has not yet been developed further. Coherent convergent beam electron diffraction can be considered a holographic technique in that the partly overlapping diffraction discs in the Fourier plane show interference patterns revealing the phases of the diffracted waves [25–28]. Methods for exploiting the available holographic information for

high-resolution structure analysis are under development.

1.8.3.4 Off-Axis Electron Holography

Off-axis techniques need a beam splitter to split an incoming wave into one going through the object and one which passes by, subsequently serving as a reference wave. Beam splitting can be achieved by amplitude division or wavefront division.

Amplitude Division

Amplitude division of electron waves is accomplished by diffraction at crystals, as applied in the electron interferometer set up by Marton et al. [29]. Matteucci et al. [30] used this principle for off-axis holography, which was recently experimentally developed further by Ru [31]. The main advantage over wavefront-division techniques is the generally lower requirement for lateral coherence.

Wavefront Division

The electron biprism developed by Möllenstedt and Düker [32] is the most successful device for electron interferometry; it uses the wavefront division principle.

A series of electron interference experiments has been conducted; for example, for:

(a) Measurement of the de-Broglie wavelength [32].
(b) Electron interference microscopy [33, 34].
(c) Phase shift by mean inner potential of thin films [35].
(d) Experimental proof of the Aharonov–Bohm effect [36].
(e) Flux quantization in superconducting hollow cylinders [37, 38].
(f) Determination of contact potentials [39].
(g) Study of lateral [40, 41] and longitudinal [11, 42] coherence.
(h) Subatomic height differences and surface potentials [43].
(i) Phase shift due to magnetization in thin films [44, 45].
(j) Sagnac effect [46].

These experiments set the basis for electron holography in that they developed the deep understanding and thorough experience in experimental handling of electron wave optics. In fact, Möllenstedt and Wahl, highly experienced in electron interferometry, conducted the first experiment on off-axis electron holography using the electron biprism as a beam splitter (Fig. 4).

Off-Axis Fresnel Holography

Möllenstedt and Wahl [47], following the Gabor idea of lensless imaging, took an off-axis Fresnel hologram without any objective lens and reconstructed the image wave by means of a laser beam. Further progress was achieved by Tomita et al. [48] using a strongly defocused objective lens for taking the hologram; they also discussed the role of defocus for the achievable resolution. Saxon [49] performed the first experiments on a posteriori correction of aberrations. Recently, Mankos et al. [50] realized off-axis Fresnel holography in a scanning transmission electron microscope (STEM) for the investigation of magnetic microstructures. In general, however, off-axis Fresnel

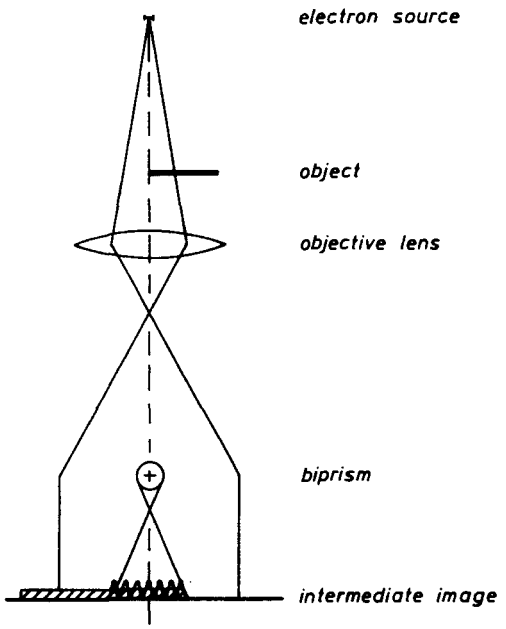


Figure 4. Taking off-axis holograms with the electron biprism. The wave to be recorded and the empty reference wave passing the biprism filament on opposite sides are superimposed to form the hologram. Interference fringes arise which are shifted aside and modulated in contrast by the phase and amplitude of the recorded wave, respectively.

holography could not even roughly reach a lateral resolution comparable to conventional electron microscopy.

Off-axis Image Plane Holography

By means of a careful study of the role of coherence for the different geometries employable for holography, Weingärtner et al. [51] showed that the highest resolution can be obtained only from holograms taken at about zero defocus, that is in the image plane of the object. Consequently, according to Bryngdahl and Lohmann [52], the interference microscopy results obtained by Möllenstedt and Buhl [33] and Faget and Fert [34] were in fact image plane holograms, although they had not been recognized as such. Interestingly, up to now image plane holography has been the most successful electron holographic technique.

1.8.4 Off-Axis Image Plane Holography

1.8.4.1 Principles

The procedure of image plane holography (i.e., taking the holograms and reconstructing the wave) was systematically developed and studied by Wahl [53]. The electron biprism inserted between the objective lens and the intermediate image plane superimposes a tilted empty reference wave of modulus 1 to the image wave

$$b(x,y) = A(x,y)\exp[\mathrm{i}\Phi(x,y)] \qquad (22)$$

In the image plane, a hologram

$$I_{\mathrm{hol}}(x,y) = 1 + A^2(x,y) + 2VA(x,y) \times \cos[2\pi R_{\mathrm{c}}x + \Phi(x,y)] \qquad (23)$$

results. The image wave is riding on a carrier frequency R_{c} which has to meet the condition $R_{\mathrm{c}} \geqslant 3R_{\max}$ where $R_{\max}$ is the maximum spatial frequency of the image wave contributing to the hologram, and V is the contrast of the hologram fringes. The hologram is recorded by means of an electron detector.

At the beginning of the procedure the reconstruction of the image wave was performed light optically. In the back-focal plane of a reconstruction lens, the Fourier transform of the hologram is

given by

$$\begin{aligned} \mathrm{FT}[I_{\mathrm{hol}}] = {} & \delta(R) + \mathrm{FT}[A^2] && \text{Center band} \\ & + \mathrm{FT}[A\exp(\mathrm{i}\Phi)] \\ & \otimes \delta(R - R_{\mathrm{c}}) && +1 \text{ Sideband} \\ & + \mathrm{FT}[A\exp(-\mathrm{i}\Phi)] \\ & \otimes \delta(R + R_{\mathrm{c}}) && -1 \text{ Sideband} \end{aligned} \tag{24}$$

Cutting out one sideband, after centering around $R = 0$ and inverse Fourier transformation, one obtains the image wave or the conjugate complex, respectively. At this stage, the wave optical tools can be applied (e.g., to extract amplitude and phase separately). Wahl demonstrated the possibilities of phase analysis by superimposing a plane wave reconstructed from another hologram, or from a second 'empty' hologram recorded on the same photographic plate ('double exposure hologram') [54]. Likewise he showed the facility for the generation of a focal series from one hologram.

In particular, at high resolution the condition $R_{\mathrm{c}} \geqslant 3R_{\max}$ poses severe problems with regard to recording the fine fringe spacing needed. Therefore, as a special development of image plane holography, the technique of phase-shifting holography is adopted from light optics [55–57]. Despite the fact that the fringes are coarser than the image details, by using at least three holograms taken at different fringe positions the image wave can be retrieved.

1.8.4.2 Further Developments

Producing the Hologram

Since the work of Wahl, the steps of recording the holograms and reconstructing the image wave have been improved considerably. The indispensable highly coherent field emission sources, which were previously used only in special experiments in electron interferometry and holography [20, 22, 41, 43, 49, 58, 59], are now generally available. Microscopes for generating holograms have been developed that are dedicated to special purposes; for example, for magnetic specimens at low ambient magnetic fields [60, 61]; equipped with a helium cold stage for superconducting specimens [62]; with a superconducting lens system (e.g., for biologic specimens) [63]; or for the needs of atomic resolution [64].

The geometry of the hologram is given by the position of the biprism along the optical axis, and by the magnification factor in the subsequent intermediate image. The fringe spacing can be selected by altering the filament voltage. The hologram width also depends on the filament voltage, and can also be varied by manipulating the distance to the subsequent intermediate image plane and hence can be set by varying the excitation of the intermediate lens. To achieve greater flexibility with regard to fringe spacing and hologram width, in addition to the usual position of the biprism in the Selected Area aperture, Hitachi offers a second port for the electron biprism between the two intermediate lenses. A rotatable biprism [65] can be very helpful to bring the hologram into line with an object detail. For high resolution holography, the axial position of the biprism should be optimized [66].

Recording the Hologram

The detector has to be examined according to the following criteria:

(a) The modulation transfer function (MTF) decreases the contrast of the hologram fringes.
(b) The output signal must be *linearly* related to the electron current density distribution; otherwise the amplitude (not the phase) is distorted, resulting in artifacts under a posteriori focusing or correction of aberrations.
(c) The detection quantum efficiency (DQE) increases the noise in the hologram.
(d) A high dynamic range enables a high dose to be collected without running into saturation effects.
(e) Geometric distortions produce artificial bending of the recorded wave.
(f) The time delay between hologram recording and reconstruction, for example by darkroom processing.

With regard to most of these criteria, the still often used photoplates are not the best choice; darkroom processing takes a lot of time, the results are only reproducible with exact processing procedures, and the correction of nonlinearity, which is indispensable for many applications, is an awkward task [67]. Modern CCD cameras [68] are much more favorable in all these respects [69, 70], except in the available pixel number limiting the recordable field of view. CCD cameras for electron microscopy are available with 1000×1000 pixels, and recently 2000×2000 pixel cameras have been marketed by Photometrics and GATAN.

Reconstruction of the Electron Object Wave

The light optical reconstruction proposed by Gabor has the advantage of high reconstruction speed. In addition, the usual techniques for phase contrast and interferometry (e.g., by means of a Mach–Zehnder interferometer [71, 72]) can readily be applied. However, in particular for quantitative analysis and the correction of aberrations, numerical reconstruction is a far superior technique. Furthermore, with the rapid development in computing power, numerical reconstruction can now be performed in almost real-time. Hybrid techniques have been developed [73] which, by means of a TV camera and a liquid crystal display (LCD), transfer the hologram to an optical bench in order to proceed very quickly with the optical techniques applicable for time-dependent phenomena. In another set-up, phase-plates for focusing and aberration correction are generated by a computer and, using an LCD arranged in the Fourier plane, conveyed to the optical bench.

For a purely numerical reconstruction, the holograms are recorded by means of a CCD camera, digitized and fed to a computer [74]. Usually, for wave optical processing, image processing software is applied (e.g., IMAGIC or SEMPER) which is equipped with user-written subroutines for the particular purpose. Special developments in hardware and software allow automated or operator-controlled reconstruction within a couple of seconds [75–77]. For very careful correction of aberrations, a special program allows the operation of a fast computer like a microscope in that arbitrary values for focus, spherical aberration, etc., are selected and, after about 1 s, the corresponding amplitude and phase images of the reconstructed wave are displayed [78]. A new approach to numerical reconstruction using neural networks is described by Heindl et al. [57].

1.8.4.3 Holography at Low and Medium Resolution

For the analysis of structures consisting of spatial frequencies smaller than about $R_{max} = 0.1\ nm^{-1}$, the aberrations need not be taken into account if the electron microscope is aligned with care. The contribution from a spherical aberration to the wave aberration $\chi(\boldsymbol{R})$ can be neglected; likewise, defocus at roughly Gauss focus does not contribute much either. Consequently, without too much effort, at low resolution $\cos\chi \approx 1$ and $\sin\chi \approx 0$ (i.e., ideal aberration-free imaging conditions can be reached), and the reconstructed wave agrees perfectly with the object wave. At medium resolution, up to about half Scherzer resolution R_{scherz}, a Gabor focus Dz_{gab} of $0.56Dz_{scherz}$ should be selected, which gives a sufficiently aberration-free object wave up to $R_{max} = 0.884R_{scherz}$ without the need to correct for aberrations [79]. Therefore, the main and most essential benefit of electron holography at low and medium resolution lies in the superb wave optical facilities for the analysis of amplitudes and phases, as shown by many applications.

Applications: Magnetic Structures

The contour lines of equal phases modulo 2π representing the phase distribution can be interpreted such that between two successive lines the magnetic flux $\Phi_0 = h/e$ is enclosed. In the case of homogeneous magnetic fields restricted to the object, the contour lines show the lines of the magnetic field $\boldsymbol{B}$ [80]. If stray fields are present, interpretation of the phase distribution in terms of field lines needs care because stray fields in the direction of the electron path sum together, and far-reaching stray fields may likewise modulate the reference wave. The power of holography for the analysis of magnetic microstructures has been shown by many examples [50, 81–83]. Certainly, the main progress in the holographic investigation of magnetic microstructures was achieved by Tonomura's group. The reader is referred to [84], where a detailed overview can be found. Typical applications are the analysis of:

(a) Magnetization in thin films and the structure of magnetic domain walls.
(b) Magnetization in and stray field around single and multiple domain particles.
(c) Stray field distribution around tiny magnets produced by microlithography.
(d) Confirmation of the Aharonov–Bohm effect at micrometer dimensions.
(e) Stray field around ferromagnetic tips [60, 85].

The results obtained with superconductors, for example the study of single fluxons and, most recently, the study of vortex lattices in niobium at different ambient magnetic field strengths (Fig. 5) [86]. The experimental results are in good agreement with the theoretical description of the phase shift produced by a vortex [87, 88], and are very impressive.

Applications: Electric Structures

The phase distribution of the reconstructed wave represents the projected potential along the electron paths. Therefore, in the case of a three-dimensional potential distribution changing strongly along the electron path, the interpretation

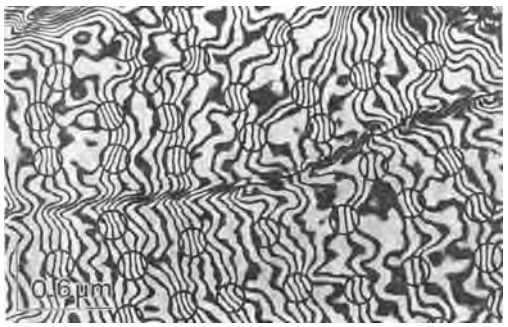


Figure 5. Phase image of flux-line lattice of niobium specimen at 4.5 K and 100*g* magnetic field. The phase contour lines represent the projected magnetic lines of force indicating the vortices at the densified locations [86]. (Courtesy of A. Tonomura and Elsevier Publishers.)

of the phase lines may be rather complicated [89]. Again, far-reaching stray fields may affect the reference wave. In the case of electric fields limited to a well-defined object area, the phase distribution correlates with the potential distribution. Typical applications of holography in this field aim at the analysis of:

(a) Stray fields around Latex spheres [90] and charged microtips [91].
(b) Potential distributions over pn junctions [92–94].
(c) Structure of silica gate oxide layers of field-effect transistors (FETs) [95].
(d) Mean inner potentials of inorganic [96] and organic [97] materials.
(e) Dynamic phase shifts in crystals [98].
(f) Inelastic mean free path evaluation of the amplitude [99].
(g) Ferroelectric domain walls [100].
(h) Morphology of nanometer-sized particles [101, 102].
(i) Cross-sections of different fullerene materials [103].
(j) Potential distributions across grain boundaries [104].
(k) Theoretical investigations of linear charged dislocations [105].
(l) Composition profiles across heterogeneous interfaces [106].

In the field of biological objects, the use of holography can obviate the need for staining [107, 108]; at present, however, due to the radiation damage problem, noise still prohibits results equivalent to those obtained with conventional microscopy from being obtained. Hopefully, the use of a superconducting objective lens [63] will solve this problem.

The analysis of surfaces by reflection holography is made possible by superimposing waves which are Bragg-reflected at different areas of a crystal surface. The reconstructed wave shows a high sensitivity of the electron phase also for surfaces with subatomic topology [109, 110].

Another special technique is convergent beam interferometry, where two selected discs of the convergent beam diffraction pattern are superimposed by means of an electron biprism [111]. This method is still under development.

1.8.4.4 Holography at High Resolution: Surmounting the Point Resolution

Beyond Gabor resolution, both $\cos(\chi)$ and $\sin(\chi)$ begin to oscillate heavily. Therefore, neither the amplitude nor the phase of the reconstructed wave can be interpreted in terms of the amplitude and phase of the object. Therefore, aberrations must be corrected during reconstruction prior to wave optical analysis [112, 113].

Correction of aberrations is accomplished by the application of a numerically generated wave transfer function $\mathrm{WTF}_{\mathrm{num}}(\boldsymbol{R})$ to the Fourier spectrum of the reconstructed image wave. Then the reconstructed object wave is given by

$$o_{\mathrm{rec}}(x,y) = \mathrm{FT}^{-1}\{\mathrm{FT}[o(x,y)] \cdot \mathrm{WTF}_{\mathrm{mic}}(\boldsymbol{R})/\mathrm{WTF}_{\mathrm{num}}(\boldsymbol{R})\} \tag{25}$$

with the wave transfer functions

$$\mathrm{WTF}_{\mathrm{mic}}(\boldsymbol{R}) = E_{\mathrm{coh}}(\boldsymbol{R}) \cdot E_{\mathrm{chrom}}(\boldsymbol{R}) \cdot \exp[-\mathrm{i}\chi_{\mathrm{mic}}(\boldsymbol{R})] \tag{26}$$

for the electron microscope and $\mathrm{WTF}_{\mathrm{num}}(\boldsymbol{R})$ numerically modeling the microscope accordingly [114].

Of course, the goal is that $o_{\mathrm{rec}}(x,y)$ approaches the object wave $o(x,y)$ as closely as possible, for instance to reach a resolution of 0.1 nm. This means that the highest possible spatial frequencies must be caught in the hologram, no additional artifacts must be introduced by the method, and the reconstruction and correction procedures have to be performed as carefully as possible. With a 100 kV electron microscope, a resolution of 0.2 nm has been achieved [115]. A thorough analysis of the holographic steps shows that a desirable resolution of 0.1 nm will be possible at an electron energy of 300 kV, if the following points are adhered to [116]; experimentally, about 0.13 nm has been reached up-to-date (Fig. 6).

Making the Hologram

The information limit R_{lim} of the electron microscope must be larger than R_{max}, the highest spatial frequency of interest.

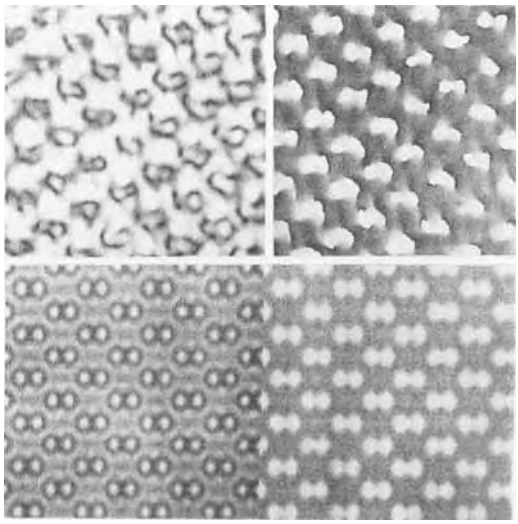


Figure 6. Object wave of silicon in the ⟨110⟩ orientation reconstructed from a 300 kV hologram. The dumbbells with a spacing of 0.136 nm can be discerned in both the amplitude (left) and phase (right) images. (Top) Holographic reconstruction; (bottom) corresponding simulated results. (From [136].)

Therefore, for the chromatic envelope function $E_{\mathrm{chrom}}(\boldsymbol{R})$ a small coefficient C_{c} of chromatic aberration and a small energy spread is needed; with a field emission gun and usual C_{c} values, at 300 kV, $E_{\mathrm{chrom}}(\boldsymbol{R})$ reaches beyond $R = 10\,\mathrm{nm}^{-1}$. However, the coherence envelope function

$$E_{\mathrm{coh}}(\boldsymbol{R}) = \exp\left\{-\left[\frac{\theta_{\mathrm{c}}}{\sqrt{\ln 2}\cdot\lambda}\,\mathrm{grad}\chi_{\mathrm{mic}}(\boldsymbol{R})\right]^2\right\} \tag{27}$$

where $\mathrm{grad}[\chi_{\mathrm{mic}}(\boldsymbol{R})]$ depends on the focus and θ_{c} is the illumination semi-aperture. With modern microscopes one usually finds, at Scherzer focus, $R_{\mathrm{lim}} \leqslant 7\,\mathrm{nm}^{-1}$, and hence holographically a resolution of about 0.1 nm cannot be reached at Scherzer focus.

Surprisingly, for high resolution the hologram must have a minimum width of

$w \geqslant 4\mathrm{PSF}$, where

$$\mathrm{PSF} = \mathrm{grad}[\chi_{\mathrm{mic}}(\boldsymbol{R})]|_{\max}/\pi \tag{28}$$

is the diameter of the PSF in the image plane of the electron microscope; $\mathrm{grad}[\chi_{\mathrm{mic}}(\boldsymbol{R})]|_{\max}$ is the maximum value of $\mathrm{grad}[\chi_{\mathrm{mic}}(\boldsymbol{R})]$ for $|\boldsymbol{R}|$ in the range $[0, R_{\max}]$. Evidently, in contrast to Scherzer focus, which optimizes the $\chi_{\mathrm{mic}}(\boldsymbol{R})$ function for phase contrast in conventional microscopy, in holography the optimum focus

$$Dz_{\mathrm{opt}} = 0.75 C_{\mathrm{s}} \left(\frac{R_{\max}}{k} \right)^2 \tag{29}$$

has to be chosen to minimize $\mathrm{grad}[\chi_{\mathrm{mic}}(\boldsymbol{R})]|_{\max}$ [79]. Then, compared to the Scherzer focus, with the PSF taking the value

$$\mathrm{PSF} = 0.5 C_{\mathrm{s}} \left(\frac{R_{\max}}{k} \right)^3 \tag{30}$$

the PSF is reduced by a factor 4; additionally, the information limit R_{lim} is increased by a factor of about 1.6 and hence exceeds the desirable threshold of $10\,\mathrm{nm}^{-1}$ [64].

Taking account of the condition $R_{\mathrm{c}} \geqslant 3R_{\max}$, at optimum focus one finds that a minimum number of fringes

$$N_{\mathrm{fringe}} = R_{\mathrm{c}} w = 6 C_{\mathrm{s}} \frac{R_{\max}^4}{k^3} \tag{31}$$

is needed for a holographic resolution $R_{\max}$. With a 300 kV microscope, typical values are $w \geqslant 20\,\mathrm{nm}$ and $N_{\mathrm{fringe}} = 600$ for a resolution of 0.1 nm.

Recording the Hologram

Because of the drawbacks of the photoplate, in particular the nonlinearity producing artificial distortions of the recorded wave, for high resolution holography the choice is a CCD camera. Here, the critical point is the available pixel number, which may limit the holographic resolution. The reason for this is that each fringe has to be sampled by at least 4 pixels [117]. Therefore the number of needed pixels is

$$N_{\mathrm{pix}} \geqslant 4 N_{\mathrm{fringe}} = 24 C_{\mathrm{s}} \frac{R_{\max}^4}{k^3} \tag{32}$$

and, by inversion, one finds

$$R_{\max} = \left[\frac{N_{\mathrm{pix}} k^3}{24 C_{\mathrm{s}}} \right]^{1/4} \tag{33}$$

Related to the Scherzer resolution

$$R_{\mathrm{scherz}} = 1.5 \left[\frac{k^3}{C_{\mathrm{s}}} \right]^{1/4} \tag{34}$$

of the microscope a holographic improvement factor of resolution

$$\frac{R_{\max}}{R_{\mathrm{scherz}}} = 0.3 N_{\mathrm{pix}}^{1/4} \tag{35}$$

results, given by the pixel number of the CCD camera. From a hologram taken with a microscope at point resolution R_{scherz} and a pixel number N_{pix}, the resolution of the reconstruction is, at best, $R_{\max}$ [12]. A different approach is discussed in [118].

Reconstruction of the Object Exit Wave by Correction of Aberrations

The carrier frequency R_{c} and orientation of the hologram fringes determine the position of the sideband in Fourier space. In general, the center of the sideband does not coincide exactly with a pixel, and thus strong streaks arise which severely hamper the cutting out of the sideband for further processing and quantitative analysis in Fourier space. Furthermore, the reconstructed wave would be tilted with respect

to the optical axis. Therefore, the sideband must be centered around the optical axis with subpixel accuracy. Since interpolation would produce severe artifacts, special methods for subpixel centering, which involve a precise determination of the hologram parameters, have been developed [78, 119].

Next, for the correction of aberrations, the wave transfer function

$$\mathrm{WTF}_{\mathrm{num}}(\boldsymbol{R}) = E_{\mathrm{num}}(\boldsymbol{R}) \cdot \exp[-\mathrm{i}\chi_{\mathrm{num}}(\boldsymbol{R})] \tag{36}$$

has to be created such that it models precisely the electron microscope:

$$\mathrm{WTF}_{\mathrm{num}}(\boldsymbol{R}) \equiv \mathrm{WTF}_{\mathrm{mic}}(\boldsymbol{R}) \tag{37}$$

delivers $\mathrm{o}_{\mathrm{rec}}(x, y) \equiv \mathrm{o}(x, y)$. Modeling the envelope function $E_{\mathrm{num}}(\boldsymbol{R})$ is not critical; however, the wave aberration $\chi_{\mathrm{num}}(\boldsymbol{R})$ must be dealt with extremely carefully.

Ideally, the resulting wave aberration describing the overall transfer of the object exit wave to the reconstructed object wave should vanish. In reality, however, a residual aberration $\delta\chi(\boldsymbol{R}) = \chi_{\mathrm{num}}(\boldsymbol{R}) - \chi_{\mathrm{mic}}(\boldsymbol{R})$ will distort the result by the transfer functions $\cos(\delta\chi)$ and $\sin(\delta\chi)$. Freedom from aberrations can be assumed if $\cos(\delta\chi) \geqslant 0.86$ and $\sin(\delta\chi) \leqslant 0.5$, that is if $|\delta\chi| \leqslant \pi/6$ holds over the entire range $[0, R_{\max}]$ [116].

In the most simple case, only spherical aberration and defocus have to be considered

$$\delta\chi(\boldsymbol{R}) = 2\pi k \left[0.25\, \delta C_{\mathrm{s}} \left(\frac{R}{k} \right)^4 - 0.5\, \delta Dz \left(\frac{R}{k} \right)^2 \right] \tag{38}$$

where δC_{s} and δDz are the imprecision occurring in the determination of the spherical aberration and defocus that are effective in the hologram; to reach $|\delta\chi| \leqslant \pi/6$, they simultaneously have to be smaller than

$$|\delta C_{\mathrm{s}}| \leqslant \frac{k^3}{6R_{\max}^4} \tag{39}$$

and

$$|\delta Dz| \leqslant \frac{k}{12R_{\max}^2} \tag{40}$$

At 300 kV and $R_{\max} = 10\,\mathrm{nm}^{-1}$, the values are $\delta C_{\mathrm{s}} = 2\,\mu\mathrm{m}$ and $\delta Dz = 0.4\,\mathrm{nm}$.

In the general case, with a comparable precision, one has also to determine:

(a) The wavenumber k.
(b) The gauging of the spatial frequencies.
(c) The two- and three-fold astigmatism [120].
(d) The axial coma.

Altogether this makes 10 parameters that are needed to model the electron microscope in the computer, all of which have to be determined for the specific hologram. However, the usual methods of diffractometry do not give sufficient accuracy, not even with sophisticated methods [121]; in addition, these data represent the microscope at the time of measurement, but not the aberrations caught in the specific hologram. For refinement, one can make use of the reconstructed wave by means of, for example:

(a) Minimizing the amplitude contrast at pure phase objects [122, 123].
(b) Minimizing the amplitude contrast in interatomic space [124].
(c) Matching the amplitude and phase of ideal crystalline areas with simulations.
(d) Evaluating the phases in the Fourier spectrum [122].

Quality of the Reconstructed Object Wave

Besides resolution, the degree of fidelity with the object exit wave and the signal-to-noise ratio properties need to be considered.

Artifacts. In holography there is a considerable risk of catching artifacts [125] from the following sources. Fresnel diffraction at the biprism produces parasitic amplitude and phase structures, which cannot easily be removed. Also, vignetting occurs due to the biprism, in that parts of the field of view are transferred as a single-sideband image with falsified amplitude and phase. Another serious artifact comes about because of the restricted field of view and the resulting broadening of reflections in reconstructed Fourier space. Large problems may arise due to geometric distortions of the hologram (e.g., from projector lenses) which falsify the phase; even worse, they restrict isoplanacy in the reconstructed wave, and hence may limit the applicability of the correction procedure and resolution. These artifacts can be avoided by taking wider holograms, and by correcting distortions by means of an empty reference hologram carrying the information needed, respectively [74].

Signal-to-Noise Properties. Correction of aberrations means a posteriori opening of the objective aperture. Therefore, the signal will be enhanced and there is some hope of detecting single atoms (e.g., gold or even oxygen with a phase shift of $2\pi/12$ and $2\pi/60$, respectively). However, noise must be accordingly low [126–129].

The phase detection limit from a hologram is given by

$$\delta\phi = \frac{\mathrm{SNR}}{V}\sqrt{\frac{2}{N'}} \tag{41}$$

The contrast $V = |\mu| \cdot V_{\mathrm{inst}} \cdot V_{\mathrm{inel}} \cdot V_{\mathrm{MTF}}$ is given by the degree of coherence, the instabilities, the inelastic interaction and the MTF of the detector. $N' = N \cdot \mathrm{DQE}$ means the number of electrons per resolved area reduced by the DQE of the detector.

Inserting

$$N = j_{\mathrm{coh}} \epsilon t / R_{\max}^2 \tag{42}$$

$$j_{\mathrm{coh}} = \frac{-2B_0 \ln(|\mu|)}{\pi k^2 w^2} \tag{43}$$

and

$$w = 2C_{\mathrm{s}}\left[\frac{R_{\max}}{k}\right]^3 \tag{44}$$

one obtains

$$\delta\phi = \frac{2\sqrt{\pi} C_{\mathrm{s}} R_{\max}^4 \mathrm{SNR}}{|\mu| \cdot V_{\mathrm{inel}} \cdot V_{\mathrm{inst}} \cdot V_{\mathrm{MTF}} \cdot k^2 \sqrt{-B_0 \ln(|\mu|) t\epsilon \cdot \mathrm{DQE}}} \tag{45}$$

Note that C_{s} is an essential limiting factor, despite the fact that it is corrected at the end. For optimum signal-to-noise ratio (SNR) one has to take into account:

(a) That $|\mu|\sqrt{-\ln(|\mu|)}$ is maximum at $|\mu| = 0.61$.
(b) The brightness B_0 of the electron gun must be a maximum.
(c) There must be a trade-off between exposure time t and V_{inst}.
(d) The need for extremely low AC magnetic stray fields, mechanical stability of biprism and the object, and vibrations and acoustic noise.
(e) That a priori correction of spherical aberration [130] could help greatly.

With special care, $\delta\phi \leqslant 2\pi/60$ at SNR = 3 and $R_{\max} = 10\,\mathrm{nm}^{-1}$ is within reach.

Wave Optical Analysis

The reconstructed object wave can be analyzed systematically and quantitatively wave optically from the amplitude and phase image. In our experience, the phase image is often much easier to interpret in that, for example, there are no half-spacings and local crystal tilt can readily be seen by asymmetric phase structures at atomic columns. Phase jumps at the extinction thickness are readily discernible. Application of a numerical selected-area aperture with a diameter down to less than 1 nm and Fourier transform yields nano-diffraction patterns for a local analysis of crystal tilt and thickness effects. Both the amplitude and the phase of Fourier spectra can be studied; there is no phase problem with holography. Selecting single reflections or groups by using an aperture allows one to display the object wave under arbitrary 'selective imaging' conditions such as bright field, dark field, or diffraction contrast, and to investigate, for example, the scattering phases for each reflection [131].

Since all this is done using one hologram, the data obtained can be mutually related in a definite way. In addition to the availability of the phases, it is a great advantage that during the analysis neither the 'microscope' or the 'specimen' change (e.g., by shift of alignment, drift, contamination, or radiation damage). Applications to the study of real crystal structures are reported in [132–134].

Outlook

With the CM30FEG-Special Tuebingen microscope dedicated to high resolution electron holography ($C_s = 1.2\,\text{mm}$, $B_0 = 5 \times 10^8\,\text{A cm}^{-2}\,\text{sr}^{-1}$, $R_{\text{scherz}} = 5\,\text{nm}^{-1}$, $R_{\text{lim}} = 10\,\text{nm}^{-1}$) and a 1000×1000 pixel CCD-camera, a holographic resolution of $R_{\text{max}} = 7.5\,\text{nm}^{-1}$ was achieved [135, 136]. After upgrading with the UltraTwin lens ($C_s = 0.62\,\text{mm}$, $R_{\text{scherz}} = 5.7\,\text{nm}^{-1}$), and a 2000×2000 pixel CCD camera yielding a resolution improvement factor of 2, a holographic resolution of $R_{\text{max}} = 10\,\text{nm}^{-1}$ seems realistic.

1.8.5 Summary

After more than 40 years of development, electron holography has reached a state which shows the high potential for microstructure analysis beyond the limits of conventional electron microscopy. This has already been proven by a wide range of applications in materials science.

At present, off-axis image plane holography is the most advanced holographic technique, at both low and high resolution. Presumably, the other techniques under development will also show special areas of applications where they will allow unique features to be observed, and new techniques will arise with the general progress in electron wave optics.

1.8.6 References

[1] W. Glaser, *Z. Phys.* **1933,** *80*, 451.
[2] W. Ehrenberg, R. E. Siday, *Proc. Phys. Soc. B* **1949,** *62*, 8.
[3] Y. Aharonov, D. Bohm, *Phys. Rev.* **1959,** *115*, 485.
[4] O. Scherzer, *Z. Phys.* **1936,** *101*, 593.
[5] O. Scherzer, *J. Appl. Phys.* **1949,** *20*, 20.
[6] D. Gabor, *Nature* **1948,** *161*, 777.
[7] D. Gabor, G. W. Stroke, D. Brumm, A. Funkhouser, A. Labeyrie, *Nature* **1965,** *208*, 1159.

[8] L. O. Heflinger, R. F. Wuerker, R. E. Brooks, *J. Appl. Phys.* **1966,** *37*, 642.
[9] O. Bryngdahl, *J. Opt. Soc. Am.* **1969,** *59*, 142.
[10] D. Gabor, *Proc. R. Soc. London* **1951,** *64*, 449.
[11] H. Schmid, PhD. Thesis, Tübingen **1985.**
[12] H. Lichte, *Ultramicroscopy* **1993,** *51*, 15.
[13] E. Leith, J. Upatnieks, *J. Opt. Soc. Am.* **1962,** *52*, 1123.
[14] J. M. Cowley, *Ultramicroscopy* **1992,** *41*, 335.
[15] M. E. Haine, T. Mulvey, *J. Opt. Soc. Am.* **1952,** *42*, 763.
[16] H.-W. Fink, H. Schmid, H. J. Kreuzer, A. Wierzbicki, *Phys. Rev. Lett.* **1991,** *67*, 1543.
[17] H. J. Kreuzer, H.-W. Fink, H. Schmid, S. Bonev, *J. Microsc.* **1995,** *178*, 191.
[18] J. C. H. Spence, X. Zhang, W. Qian, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 267.
[19] A. Tonomura, A. Fukuhara, H. Watanabe, T. Komoda, *Jpn. J. Appl. Phys.* **1968,** *7*, 295.
[20] N. Bonnet, M. Troyon, P. Gallion, in *Proc. 9th Int. Cong. Electron Microscopy*, Microscopical Society of Canada, Toronto **1978,** vol. I, pp. 222–223.
[21] T. Matsumoto, T. Tanji, A. Tonomura, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 249.
[22] J. Munch, *Optik* **1975,** *43*, 79.
[23] T. Matsumoto, T. Tanji, A. Tonomura, *Optik* **1995,** *100*, 71.
[24] R. Lauer, *Optik* **1984,** *66*, 159.
[25] J. Cowley, in *Electron Diffraction Techniques* (Ed.: J. M. Cowley), Oxford Science Publications, Oxford **1992,** vol. I, 439.
[26] J. W. Steeds, P. A. Midgley, P. Spellward, R. Vincent, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 277.
[27] J. A. Lin, J. M. Cowley, in *Proc. 43rd Meeting of the Electron Microscopical Society of America*, Electron Microscopical society of America, **1985,** 136.
[28] J. M. Cowley, M. A. Gribelyuk, *MSA Bull.* **1994,** *24*, 438.
[29] L. Marton, *Phys. Rev.* **1952,** *85*, 1057.
[30] G. Matteucci, G. F. Missiroli, G. Pozzi, *Ultramicroscopy* **1981,** *8*, 109.
[31] Q. Ru, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 343.
[32] G. Möllenstedt, H. Düker, *Z. Phys.* **1956,** *145*, 377.
[33] R. Buhl, *Z. Phys.* **1959,** *155*, 395–412.
[34] J. Faget, Ch. Fert, *C. Phys.* **1987,** *11*, 285.
[35] M. Keller, *Z. Phys.* **1961,** *164*, 274.
[36] W. Bayh, *Z. Phys.* **1962,** *169*, 492.
[37] H. Wahl, *Optik* **1970,** *30*, 577.
[38] B. Lischke, *Z. Phys.* **1970,** *239*, 360.
[39] E. Krimmel, G. Möllenstedt, W. Rothemund, *Appl. Phys. Lett.* **1964,** *5*, 209.
[40] T. Hibi, S. Takahashi, *Z. Phys.* **1969,** *3*, 132.
[41] R. Speidel, D. Kurz, *Optik* **1977,** *49*, 173.
[42] F. Lenz, G. Wohland, *Optik* **1984,** *67*, 315.
[43] H. Lichte, G. Möllenstedt, *J. Phys. E: Sci. Instrum.* **1979,** *12*, 941.
[44] A. Tonomura, *Jpn. J. Appl. Phys.* **1972,** *11*, 493.
[45] G. Pozzi, G. F. Missiroli, *J. Microsc.* **1973,** *18*, 103.
[46] F. Hasselbach, M. Nicklaus, *Phys. Rev. A* **1993,** *48*, 143.
[47] G. Möllenstedt, H. Wahl, *Naturwissenschaften* **1968,** *55*, 340.
[48] H. Tomita, T. Matsuda, T. Komoda, *Jpn. J. Appl. Phys.* **1972,** *11*, 143.
[49] G. Saxon, *Optik* **1972,** *35*, 359.
[50] M. Mankos, P. de Haan, V. Kambersky, G. Matteucci, M. R. McCartney, Z. Yang, M. R. Scheinfein, J. M. Cowley, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 329.
[51] I. Weingärtner, W. Mirandé, E. Menzel, *Optik* **1969,** *30*, 318.
[52] O. Bryngdahl, A. W. Lohmann, *J. Op. Soc. Am.* **1968,** *58*, 141.
[53] H. Wahl, Thesis, Tübingen **1975.**
[54] H. Wahl, *Optik* **1974,** *39*, 585.
[55] W. D. Rau, H. Lichte, K.-H. Herrmann, *Optik* **1989,** *83* (Suppl 4), 79.
[56] Q. Ru, G. Lai, K. Aoyama, J. Endo, A. Tonomura, *Ultramicroscopy* **1994,** *55*, 209.
[57] E. Heindl, W. D. Rau, H. Lichte, *Ultramicroscopy*, in press.
[58] W. Brünger, *Z. Phys.* **1972,** *250*, 263.
[59] A. Tonomura, T. Matsuda, T. Komoda, *Jpn. J. Appl. Phys.* **1978,** *17*, 1137.
[60] B. Frost, Thesis, Tübingen **1993.**
[61] T. Hirayama, J. Chen, Q. Ru, K. Ishizuka, T. Tanji, A. Tonomura, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 145.
[62] A. Tonomura, N. Osakabe, T. Matsuda, T. Kawasaki, J. Endo, S. Yano, H. Yamada, *Phys. Rev. Lett.* **1986,** *56*, 792.
[63] U. Weierstall, H. Lichte, *Ultramicroscopy*, in press.
[64] H. Lichte, P. Kessler, F. Lenz, W.-D. Rau, *Ultramicroscopy* **1993,** *52*, 575.
[65] K. Yada, K. Shibata, T. Hibi, *J. Elec. Microsc.* **1973,** *22*, 223.

[66] H. Lichte, *Ultramicroscopy*, in press.
[67] E. Völkl, F. Lenz, Q. Fu, H. Lichte, *Ultramicroscopy* **1994,** *55*, 75.
[68] J. H. C. Spence, J. M. Zuo, *Rev. Sci. Instrum.* **1988,** *59*, 2102.
[69] I. Daberkow, K.-H. Herrmann, Libin Liu, W. D. Rau, *Ultramicroscopy* **1991,** *38*, 215.
[70] O. L. Krivanek, P. E. Mooney, *Ultramicroscopy* **1993,** *49*, 95.
[71] A. Tonomura, J. Endo, T. Matsuda, *Optik* **1979,** *53*, 143.
[72] K.-J. Hanszen, *Adv. Electron. Electron Phys.* **1982,** *59*, 1.
[73] J. Chen, T. Hirayama, G. Lai, T. Tanji, K. Ishizuka, A. Tonomura, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 81.
[74] W. D. Rau, H. Lichte, E. Völkl, U. Weierstall, *J. Comput. Assist. Microsc.* **1991,** *3*, 51.
[75] W. D. Rau, *MSA Bull.* **1994,** *24*, 459.
[76] E. Voelkl, L. F. Allard, B. Frost, *J. Microsc.*, **1995,** *180*, 39.
[77] M. Lehmann, E. Völkl, F. Lenz, *Ultramicroscopy* **1994,** *54*, 335.
[78] M. Lehmann, H. Lichte, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 69.
[79] H. Lichte, *Ultramicroscopy* **1991,** *38*, 13.
[80] H. Wahl, B. Lau, *Optik* **1979,** *54*, 27.
[81] B. Lau, G. Pozzi, *Optik* **1978,** *51*, 287.
[82] G. Matteucci, G. F. Missirolli, J. W. Chen, G. Pozzi, *Appl. Phys. Lett.* **1988,** *52*, 176.
[83] G. Matteucci, M. Muccini, *Ultramicroscopy* **1994,** *53*, 19.
[84] A. Tonomura, *Springer Ser. Opt. Sci.* **1993,** *70*.
[85] G. Matteucci, M. Muccini, U. Hartmann, *Appl. Phys. Lett.* **1993,** *62*, 1839.
[86] J. E. Bonevich, K. Harada, H. Kasai, T. Matsuda, T. Yoshida, G. Pozzi, A. Tonomura, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 135.
[87] A. Migliori, G. Pozzi, *Ultramicroscopy* **1992,** *41*, 169.
[88] A. Migliori, G. Pozzi, A. Tonomura, *Ultramicroscopy* **1993,** *49*, 87.
[89] J. W. Chen, G. Matteucci, A. Migliori, E. Nichelatti, G. Pozzi, M. Vanzi, *Phys. Rev. A* **1989,** *40*, 3136.
[90] B. G. Frost, L. F. Allard, E. Völkl, D. C. Joy, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 169.
[91] G. Matteucci, G. F. Missiroli, M. Muccini, G. Pozzi, *Ultramicroscopy* **1992,** *45*, 77.
[92] S. Frabboni, G. Matteucci, G. Pozzi, M. Vanzi, *Phys. Rev. Lett.* **1985,** *55*, 2196.
[93] C. Capiluppi, A. Migliori, G. Pozzi, *J. Microsc. Microanal. Microstruct.*, in press.
[94] M. R. McCartney, B. Frost, R. Hull, M. R. Scheinfein, D. J. Smith, E. Voelkl, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 189.
[95] W.-D. Rau, F. H. Baumann, P. K. Roy, J. A. Rentschler, A. Ourmazd, *Appl. Phys. Lett.*, in press.
[96] M. Gajdardziska-Josifovska, M. R. McCartney, W. J. de Ruijter, D. J. Smith, J. K. Weiss, J. M. Zuo, *Ultramicroscopy* **1993,** *50*, 285.
[97] M. Libera, J. Ott, Y. C. Wang, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 231.
[98] K.-J. Hanszen, *J. Phys. D: Appl. Phys.* **1986,** *19*, 373.
[99] M. R. McCartney, M. Gajdardziska-Josifovska, *Ultramicroscopy* **1994,** *53*, 283.
[100] Xiao Zhang, D. C. Joy, Y. Zhang, T. Hashimoto, L. Allard, T. A. Nolan, *Ultramicroscopy* **1993,** *51*, 21.
[101] E. Völkl, L. F. Allard, B. Frost, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 103.
[102] A. K. Datye, D. S. Kalakkad, E. Völkl, L. F. Allard, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 199.
[103] L. F. Allard, E. Völkl, S. Subramoney, R. S. Ruoff, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 219.
[104] X. Lin, V. Ravikumar, R. Rodrigues, N. Wilcox, V. Dravid, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 209.
[105] D. Cavalcoli, G. Matteucci, M. Muccini, *Ultramicroscopy* **1995,** *57*, 385.
[106] J. K. Weiss, W. J. der Ruijter, M. Gaijdarziska-Josifovska, M. R. McCartney, D. J. Smith, *Ultramicroscopy* **1993,** *50*, 301.
[107] H. Lichte, U. Weierstall, *Inst. Phys. Conf. Ser. 93* **1988,** *3*, 325.
[108] K. Aoyama, G. Lai, Q. Ru, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 239.
[109] N. Osakabe, T. Matsuda, J. Endo, A. Tonomura, *Ultramicroscopy* **1993,** *48*, 483.

[110] H. Banzhof, K.-H. Herrmann, *Ultramicroscopy* **1993,** *48*, 475.
[111] R. A. Herring, G. Pozzi, in *Electron Holography* (Eds.: A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy, Y. A. Ono), Elsevier, Amsterdam **1995,** 287.
[112] H. Lichte, *Ultramicroscopy* **1986,** *20*, 293.
[113] H. Lichte, *Adv. Opt. Electron Microsc.* **1991,** *12*, 25.
[114] Q. Fu, H. Lichte, E. Völkl, *Phys. Rev. Lett.* **1991,** *67*, 2319.
[115] A. Harscher, G. Lang, H. Lichte, *Ultramicroscopy* **1995,** *58*, 79.
[116] H. Lichte, *Ultramicroscopy* **1992,** *47*, 223.
[117] F. Lenz, E. Völkl, Stochastic limitations to phase and contrast determination in electron holography, *Proc. XIIth Int. Cong. Electron Microscopy*, San Francisco Press, Inc., Seattle **1990,** 228.
[118] K. Ishizuka, *Ultramicroscopy* **1994,** *53*, 297.
[119] E. Völkl, L. F. Allard, *MSA Bull.* **1994,** *24*, 466.
[120] O. L. Krivanek, *Ultramicroscopy* **1994,** *55*, 419.
[121] D. Typke, K. Dierksen, *Optik* **1995,** *99*, 155.
[122] Q. Fu, H. Lichte, *J. Microsc.* **1995,** *179*, 112.
[123] K. Ishizuka, T. Tanji, A. Tonomura, T. Ohno, Y. Murayama, *Ultramicroscopy* **1994,** *53*, 361.
[124] M. A. Gribelyuk, J. M. Cowley, *Ultramicroscopy* **1993,** *50*, 29.
[125] H. Lichte, D. Geiger, A. Harscher, E. Heindl, M. Lehmann, D. Malamidis, A. Orchowski, W. D. Rau, *Ultramicroscopy*, in press.
[126] H. Lichte, K.-H. Herrmann, F. Lenz, *Optik* **1987,** *77*, 135.
[127] F. Lenz, *Optik* **1988,** *79*, 13.
[128] W. J. de Ruijter, J. K. Weiss, *Ultramicroscopy* **1993,** *50*, 269.
[129] A. Harscher, H. Lichte, *Ultramicroscopy*, in press.
[130] M. Haider, G. Braunshausen, E. Schwan, *Optik* **1995,** *99*, 167.
[131] H. Lichte, E. Völkl, K. Scheerschmidt, *Ultramicroscopy* **1992,** *47*, 231.
[132] E. Schweda, J. Strähle, H. Lichte, E. Völkl, *Ultramicroscopy* **1994,** *53*, 53.
[133] T. Tanji, K. Ishizuka, *MSA Bull.* **1994,** *24*, 494.
[134] A. Orchowski, H. Lichte, *Ultramicroscopy*, in press.
[135] H. Lichte, W.-D. Rau, *Ultramicroscopy* **1994,** *54*, 310.
[136] A. Orchowski, W. D. Rau, H. Lichte, *Phys. Rev. Lett.* **1995,** *74*, 399.

General Reading

Chapter I.1/2

L. Baumann, O. Leeder, *Einführung in die Auflichtmikroskopie*, Dt. Verlag f. Grundstoffindustrie, Leipzig, **1991**.

H. Beyer in: *Handbuch der Mikroskopie*, 3rd ed. (Ed.: H. Riesenberg), Technik, Berlin, **1988**, Chap. 1.

J. R. Craig, D. J. Vaughan, *Ore Microscopy and Ore Petrography*, New York-Chichester-Brisbane, Wiley Interscience Publication, John Wiley and Sons, **1981**.

H. Freund (Ed.), *Handbuch der Mikroskopie in der Technik*, Umschau-Verlag, Frankfurt, **1968**, Vol. I-IV.

K. Michel, *Die Grundzüge der Theorie des Mikroskops in elementarer Darstellung*, 2nd ed., Wiss. Verlagsges., Stuttgart, **1964**.

G. Petzow, *Metallographisches, Keramographisches und Plastographisches Ätzen*, Borntträger, Berlin, **1994**.

H. Pichler, C. Schmitt-Riegraf, *Gesteinsbildende Mineralien im Dünnschliff*, Ferdinand Enke, Stuttgart, **1987**.

H. Schneiderhöhn, *Erzmikroskopisches Praktikum*, E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, **1952**.

Chapter I.3

T. Wilson, *Confocal Microscopy*, Academic Press, London, **1990**.

P. Dhamelincourt et al., *Laser Raman Molecular Microprobe (MOLE)*, Anal. Chem., **1979**, *51*, 414A.

D. J. Gardiner et. al. (Eds.), *Practical Raman Spectroscopy*, Springer-Verlag, Berlin, **1989**.

J. Barbillat et al., *Raman Confocal Microprobing, Imaging and Fibre-optic Remote Sensing: a Further Step in Molecular Analysis*, J. Raman Spectrosc., **1994**, *25*, 3.

R. J. H. Clark and R. E. Hester (Eds.), *Raman Microprobe Spectroscopy, G. J. Rosasco in "Advances in Infrared and Raman Spectroscopy"*, Heyden & Son, London, **1980**, Vol. 7, pp. 223-282.

D. A. Long, *Raman Spectroscopy*, McGraw-Hill, New York, **1977**.

Chapter I.4

J. B. Pawley (Ed.), *Handbook of Biological Confocal Microscopy*, 2nd edn., Plenum Press, New York, **1995**.

T. Wilson, C. J. R. Sheppard, *Theory and Practice of Scanning Optical Microscopy*, Academic Press, London, **1984**.

P. C. Cheng, T. H. Lin, W. L. Wu, J. L. Wu (Eds.), *Computer-Assisted Multidimensional Microscopies*, Springer, New York, **1994**.

C. J. R. Sheppard in: *Current Trends in Optics* (Ed.: J. C. Dainty), Academic Press, London, **1994**, 225-231.

T. Wilson in: *Microanalysis of Solids* (Eds.: B.G. Yacobi, D. B. Holt, L. L. Kazmerski), Plenum Press, New York, **1994**, 219-232.

T. Wilson (Ed.), *Confocal Microscopy*, Academic Press, London, **1990**.

Chapter I.5

J. M. Vigoureux, C. Girard, and D. Courjon. General principles of scanning tunneling optical microscopy. *Optics Letters*, **1989**, *14*, 1039.

D. W. Pohl. Scanning near-field optical microscopy. *Advances in optical and electron Microscopy*, 12, **1991**.

D. Pohl and D. Courjon, (Eds.). *Near Field Optics*, volume 242 of *NATO Series Series E: Applied Sciences*. Kluwer Academic Publishers, Dordrecht, the Netherlands, **1993**.

H. Heinzelmann and D. W. Pohl, Scanning near-field optical microscopy. *Appl. Phys. A*, **1994**, *59*, 89.

D. Courjon and C. Bainier. Near field microscopy and near field optics. *Report on Progress in Physics*, **1994**, *57*, 989.

Chapter I.6

M. W. Mackenzie (Ed.), *Advances in Applied Fourier Transform Infrared Spectroscopy*, John Wiley and Sons, Chichester, **1988**.

P. R. Griffiths, *Chemical Infrared Fourier Transform Spectroscopy*, John Wiley and Sons, NY, **1975**.

H. H. Perkampus, *Encyclopedia of Spectroscopy*, VCH, Weinheim, **1995**.
G. Socrates, *Infrared Characteristic Group Frequencies*, John Wiley and Sons, NY, **1994**.
R. G. Messerschmidt, M. A. Harthcock Editors, *Infrared Microspectroscopy: Theory and Applications*, M. Dekker, NY, **1988**.
M. Diem, *Introduction to Modern Vibrational Spectroscopy*, Wiley Interscience, NY, **1993**.
J. D. Graybeal, *Molecular Spectroscopy*, McGraw-Hill, NY, **1988**.
E. B. Wilson, J. C. Decius, P. C. Cross, *Molecular Vibrations*, McGraw-Hill, NY, **1955**.
W. L. Wolfe, G. J. Zissis (Eds.), *The Infrared Handbook*, Office of Naval Research, Washington DC, **1978**.
L. J. Bellamy, 3rd ed., *The Infrared Spectra of Complex Molecules*, Chapman and Hall, London, **1975**.

Chapter II.1

V. V. Aristov, A. I. Erko (Eds.), *X-Ray Microscopy IV*, Bogorodskii Pechatnik, Chernogolovka, Russia, **1994**.
P. C. Cheng, G. J. Jan (Eds.), *X-Ray Microscopy-Instrumentation and Biological Applications*, Springer, Berlin, **1987**.
J. Kirz, C. Jacobsen, and M. Howells, *Soft X-Ray Microscopes and Their Biological Applications*, Q. Rev. Biophys., **1995**, *28*, 33.
A. G. Michette, G. R. Morrison, and C. J. Buckley (Eds.), *X-Ray Microscopy III*, Springer, Berlin, **1992**.
D. Sayre, M. Howells, and J. Kirz (Eds.), *X-Ray Microscopy II*, Springer, Berlin, **1988**.
G. Schmahl, D. Rudolph (Eds.), *X-Ray Microscopy*, Springer, Berlin, **1984**.
G. Schmahl, P. C. Cheng, *X-Ray Microscopy, in Handbook on Synchrotron Radiation, Vol. 4* (Eds.: S. Ebashi et al.), Chap. 15, North Holland, Amsterdam, **1991**, 481.
G. Schmahl, D. Rudolph, B. Niemann, P. Guttmann, J. Thieme, and Gerd Schneider, *Röntgenmikroskopie*, Naturwissenschaften, **1996**, *83*, 61.

Chapter II.2

B. K. Agarwal, *X-Ray Spectroscopy*, Springer Series in Optical Sciences, Vol. 15, **1979**, Springer-Verlag, Berlin.
D. E. Eastman (Ed.), *Handbook on Synchrotron Radiation*, Recommended Chapters 1, 2, 12, Vol. 1, **1983**, North Holland Publishing Company, Amsterdam.
S. Ebarshi, M. Koch, E. Rubinstein (Eds.), *Handbook on Synchrotron Radiation*, Recommended Chapter 11, Vol. 4, **1991**, North Holland Publishing Company, Amsterdam.
H. Winick, S. Doniach (Eds.), *Synchrotron Radiation Research*, Plenum Press, New York, **1980**. Ch. 8.
A. G. Michette, G. R. Morrison, C. J. Buckley (Eds.), *X-Ray Microscopy III*, Proceeding of the third international conference, London, Sept. 3-7, **1990**, Springer Series in Optical Sciences, Berlin, **1990**.
R. V. Ely (Ed.), *Microfocal Radiography*, Academic Press, London, **1980**.
J. Kirz, H. Rarback, *Soft X-Ray Microscopes*, Rev. Sci. Instrum. *56*, **1985**, 1-13.
J. Cazaux, D. Mouze, *Microradiographies et Microanalyses X sans éléments focalisants*, J. Microsc. Spectrosc. Electron., *11*, **1986**, 359-378.
D. Joyeux, *Microscopie en rayons X-mous: techniques et performances*, Annales de Physiques, Colloques n^0 1, Supplément au n^0 3, Vol. 17, Juin **1992**.
D. Sayre, H. N. Chapman, *X-Ray Microscopy*, Acta Cryst. (**1995**) *A 51*, 237.
A. Michette, *X-Ray Microscopy*, Rep. Prog. Phys. *51*, **1988**, 1525.
J. Cazaux, X. Thomas, D. Mouze, *Microscopie et Microtomographie X de Laboratoire*, J. Micros. Spect. Electron. *14*, **1989**, 263.

Chapter II.3

H. H. Barrett, W. Swindell, *Radiological Imaging*, Academic Press, NY, **1981**.
G. T. Herman, *Image reconstruction from projections*, Academic Press, NY, **1980**. *Topics in Applied Physics*, Springer Verlag, Berlin, **1979**, 32.
A. K. Jain, *Fundamentals of digital image Processing*, Prentice-Hall Intern. Inc., Englewood, cliff, NJ 07632 USA, **1989**.
IEEE Trans. on Medical Imaging, **1993**, 12.
C. L. Morgan, *Basic Principles of Computed Tomography*, University Park Press, Baltimore, USA, **1983**.
Proceedings IEEE, **1983**, Special Issue n^0 3 (Computerized tomography).

J. C. Russ, *Computed assisted Microscopy*, Plenum Press, NY, **1990**, *The image processing Handbook*, 2nd Ed. CRC Press Boca Raton, USA, **1995**.

Chapter II.4

P. Hariharan, *Optical Holography*, Cambridge Univ. Press, **1984**.

M. Francon, *Holographie*, Masson Paris, **1969** (in french, an american translation should exist).

R. T. Collier, C. B. Burckhardt and L. H. Lin, *Optical Holography*, Academic Press, New York, **1971**.

A. G. Michette, G. R. Morrison, C. T. Buckley, Eds., *X-Ray Microscopy III*, Springer series, Vol. 67, Springer, Berlin **1992**.

D. Sayre, M. Howells, T. Kirz and H. Razback, Eds., *X-Ray Microscopy II*, Springer series, Vol. 56, Springer, Berlin **1988**.

Chapter II.5

S. Amelinckx, R. Gevers, J. Van Landuyt, ed., *Diffraction and Imaging Techniques in Material Science*. 2nd edition, North Holland, Amsterdam, **1978**.

G. E. Bacon, *Neutron Diffraction*. 3rd edition, Clarendon Press, Oxford, **1975**.

J. Baruchel, J. L. Hodeau, M. S. Lehmann, J. R. Regnard, C. Schlenker, ed., *Neutron and Synchrotron Radiation for Condensed Matter Studies*, HERCULES (Higher European Research Course for Uses of Large Experimental Systems), Grenoble, Les Éditions de Physique, Orsay and Springer Verlag, Berlin. Vol. 1: Theory, Instruments and Methods, **1993**; Vol. 2: Application to Solid State Physics and Chemistry, **1994**; Vol. 3: Application to Soft Condensed Matter and Biology, **1994**.

A. Guinier, *X-Ray Diffraction in Crystals, Imperfect Crystals, and Amorphous Bodies*, Dover Publications, **1994**.

L. H. Schwartz, J. B. Cohen, *Diffraction from Materials*, 2nd ed., Springer, Berlin, **1987**.

B. K. Tanner, *X-Ray Diffraction Topography*, Pergamon Press, Oxford, **1976**.

B. K. Tanner, D. K. Bowen, ed., *Characterization of Crystal Growth Defects by X-Ray Methods*, NATO ASI Series, Plenum, New York, **1980**.

Chapter IV.1.1.1

R. W. Cahn, P. Haasen, E. J. Kramer (Eds.), *Materials Science and Technology*, Vol. 2A, VCH Weinheim, New York, Basel, Cambridge, **1992**.

S. Amelinckx, *The Direct Observation of Dislocations*, Supplement 6 of "Solid State Physics" (Eds. F. Seitz and D. Turnbull), Academic Press, New York and London, **1964**.

S. Amelinckx, R. Gevers and J. Van Landuyt (Eds.), *Diffraction Techniques in Materials Science*, North Holland, Amsterdam, New York, Oxford, **1978**.

A. K. Head, P. Humble, L. M. Clarebrough, A. T. Morton and C. T. Forwood, *Computed Electron Micrographs and Object Identification*, Volume 7 of "Defects in Crystalline Solids" (Eds. S. Amelinckx, R. Gevers and J. Nihoul), North Holland, Amsterdam and London, **1973**.

J. M. Cowley, *Diffraction Physics*, North Holland/American Elsevier, **1975**.

J. C. H. Spence, *Experimental High Resolution Electron Microscopy*, Monographs on the Phyics and Chemistry of Materials, Oxford Science Publications, **1981**.

P. B. Hirsch, A. Howie, R. B. Nicholson, D. W. Pashley, M. J. Whelan, *Electron Microscopy of Thin Crystals*, Butterworths, London, **1965**.

J. M. Cowley (Ed.), *Electron Diffraction Techniques*, International Union of Crystallography, IUCr Monographs on Crystallography, Oxford Science Publications, **1993**.

G. Thomas, *Transmission Electron Microscopy of Metals*, John Wiley and Sons Inc., **1962**.

Chapter IV.1.1.2

S. Amelinckx, R. Gevers, J. Van Landuyt (Eds.), *Diffraction and Imaging Techniques in Material Science*, North Holland, Amsterdam, **1970**, **1978**.

H. Bethge, J. Heydenreich (Eds.), *Elektronenmikroskopie in der Festkörperphysik*, Springer, Berlin, **1982**.

A. K. Head, P. Humble, L. M. Clarebrough, A. J. Morton, G. T. Forwood, in: *Computed Electron Micrographs and Defect Identification*, Vol. 7 (Eds.: S. Amelinckx, R. Gevers, J. Nihoul).

P. B. Hirsch, R. B. Nicholson, A. Howie, D. W. Pashley, M. J. Whelan, *Electron Microscopy of Thin Crystals*, Butterworths, London, **1965**.

F. R. N. Nabarro (Ed.), *Dislocation in Solids*, North Holland, Amsterdam, **1979**.
G. Thomas, *Transmission Electron Microscopy of Metals*, Wiley, New York, **1962**.

Chapter IV.1.2

J. M. Cowley, *Reflection Electron Microscopy*, in: *Surface and Interface by Electron Optical Methods*, A. Howie and U. Valdre, (Eds.), Plenum Press, New York, **1988**, pp. 127-158.
J. M. Cowley, *Special Electron Microscopy Techniques*, Chapter 7 of *Physical Methods of Chemistry*, Vol. IV, *Microscopy*, Wiley, New York, **1991**, pp. 239-284.
J. M. Cowley, *Electron Microscopy*, Chapter 11 of *The Handbook of Surface Imaging and Visualization*, A. T. Hubbard (Ed.), CRC Press, Boca Raton, **1995**, pp. 131-155.
Tung Hsu (Guest Editor), *Microsc. Res: Tech.*, **1992**, *20*, 317.
P. K. Larson and P. J. Dobson, (Eds.), *Reflection High-Energy Electron Diffraction and Reflection Imaging of Surfaces*, Plenum Press, New York, **1988**.
G. Lehmpfuhl and Y. Uchida, *Observation of surface cystallography by reflection electron microscopy*, Surface Sci., **1990**, *235*, 295.
J. Liu and J. M. Cowley, *Scanning reflection electron microscopy and associated techniques for surface studies*. Ultramicros. **1993**, *48*, 381.
Z. L. Wang, *Electron reflection, diffraction and imaging of bulk crystal surfaces in TEM and STEM*, Rep. Prog. Phys., **1993**, *56*.
K. Yagi, *Reflection electron microscopy*, J. Appl. Crystallogr., **1987**, *20*, 147.
K. Yagi, *RHEED and REM*, in *Electron Diffraction Techniques*, Vol. 2, J. M. Cowley, Ed., Oxford University Press, Oxford, **1993**, pp. 260-308.

Chapter IV.1.4

H. Fujita, *In Situ Deformation by high Voltage Electron Microscopy*, Proceedings of the 9th International Congress on Electron Microscopy, Toronto, Canada, **1978**, III, 335.
T. Imura, H. Fujita (Eds.), *High Voltage Electron Microscopy*, Special Issue of Proceedings of the 8th International Congress on Electron Microscopy, Kyoto, Japan, **1986**.
K. Ura (Ed-in-chief), *Electron Microscopy of Japan*, Present and Future, The 40th Anniversary of the Japanese Society of Electron Microscopy, **1989**.
S. Nagakura (Ed.), *Special Issue on Characterization of Advanced Materials by the High Resolution Electron Microscopy and Analytical Electron Microscopy Materials Trans.* J/M, **1990**, 31, 513.

Chapter IV.1.5

J. M. Cowley (Ed.), *Electron Diffraction Techniques*, Vol. 1, Oxford University Press, Oxford, **1992**.
J. C. H. Spence and J. M. Zuo, *Electron Microdiffraction*, Plenum Press, New York, **1992**.
J. Electron Microsc. Tech., **1989**, *13*, (1) and (2).
J. W. Steeds, *Introduction to Analytical Electron Microscopy* (Eds. J. J. Hren, J. I. Goldstein and D. C. Joy), Plenum Press, New York, **1979**, 387.
M. Tanaka, M. Terauchi, K. Tsuda, *Convergent-Beam Electron Diffraction*, Vols. I, II and III, Jeol Ltd., Tokyo, **1994**.
M. H. Loretto, *Electron Beam Analysis of Materials*, Chapman and Hall, London, **1994**.
P. B. Hirsch, A. Howie, R. B. Nicholson, D. W. Pashley and M. J. Whelan, *Electron Microscopy of Thin Crystals*, Robert Krieger, Florida, **1977**.

Chapter IV.1.6

A. Howie, U. Valdre (Ed.), *Surface and Interface Characterization by Electron Optical Methods*, Plenum Press, New York, **1988**.
A. T. Hubbard (Ed.), *Handbook of Surface Imaging and Visualization*, CRC Press, Boca Raton, **1995**.
R. Rosei (Ed.), *Chemical, Structural and Electronic Analysis of Heterogeneous Surfaces on Nanometer Scale*, Kluwer Acad. Publ., Dordrecht, **1996**.
V. Veneklasen, *Rev. Sci. Instrum.*, **1992**, *63*, 5513.
E. Bauer, *Rep. Prog. Phys.*, **1994**, *57*, 895.

Chapter IV.1.7

S. Chikazumi, *Physics of Magnetism*, Wiley, **1978**.

J. P. Jakubovics, *Magnetism and Magnetic Materials*, 2nd edition; The Institute of Metals, **1987**.
C. Kittel, *Introduction to Solid State Physics*, Wiley, **1986**.

Chapter IV.1.8

E. Hecht, *Optics*, Addison Wesley Publishing Company, 2nd edition, **1987**.
R. J. Collier, C. B. Burckhardt and L. L. Lin, *Optical Holography*, Academic Press, **1971**.
G. O. Reynolds, J. B. DeVelis, G. B. Parrent Jr. and B. J. Thompson, *The New Physical Optics Notebook: Tutorials in Fourier Optics*, SPIE The International Society for Optical Engineering and American Institute of Physics, SPIE Optical Engineering Press, **1989**.
A. Tonomura, *Electron Holography*, Springer Series in Optical Sciences, Vol. 70, **1993**.
A. Tonomura, L. F. Allard, G. Pozzi, D. C. Joy and Y. A. Ono (Ed.), *Electron Holography*, Elsevier, Amsterdam, **1995**.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

List of Symbols and Abbreviations

Symbol	Designation	Chapter
$\otimes$	convolution operation	III.1
$2V$	angle of bisectrix; optical angle; angle between the optical axes	I.1/2
$2V_{obs}$	observed angle of bisectrix	I.1/2
$2V_{red}$, $2V_{violett}$	angles of acute bisectrix for the colors red and violet	I.1/2
$2V_z$	angle of bisectrix along the principal Z-axis of the indicatrix	I.1/2
A	amplitude	I.1/2
A	area of the wigner-Seitz cell	IV.1.5
A	Parameter for modelling background	IV.1.3
a	attenuation	III.1
$A(\boldsymbol{g})$	physical aperture function	IV.1.1.2
$A(\boldsymbol{h})$	scattered amplitude for scattering vector $\boldsymbol{h}$	IV.1.1
$a(\theta)$	aperture function	IV.1.7
$\boldsymbol{A}(\boldsymbol{r})$	magnetic potential field	IV.1.8
$a(x)$	aperture function	IV.1.1.2
$A(x, y)$	amplitude of a beam scattered by a column at (x, y)	IV.1.1
$a(x, y)$	amplitude of electron wave	IV.1.8
a, b, c	axes of a crystal system	I.1/2
A_0	initial amplitude after transmission	I.1/2
$\mathscr{A}_0$	amplitude of the plane wave	II.1
A_0, A_1, A_2, A_m	reflection coefficients	III.1
A_1	lateral amplification	I.1/2
$\mathscr{A}_1$	amplitude of the transmitted wave	II.1
$A_d(\theta)$	amplitude in back focal plane	IV.1.7
$A_i(\xi)$	Fourier transform of $A_d(\theta)f(\theta)$	IV.1.7
A_n	numerical aperture	I.1/2
A_{obj}	amplification of objektive	I.1/2
A_{pro}	amplification of projective	I.1/2
A_s	size asymmetry of pulse	III.1
A_t	total amplification	I.1/2
A_t	time asymmetry of pulse	III.1
B	constant	II.3
$\boldsymbol{B}$	magnetic field	IV.1.8
$\boldsymbol{b}$	Burger' vector	IV.1.5, IV.1.2, II.5, IV.1.1
b	image vector	IV.1.8
$B(\Delta, \boldsymbol{g})$	damping envelope due to defocus spread	IV.1.1.2
b^*	the conjugate complex of b	IV.1.8
$B, \boldsymbol{B}$	magnetic induction	IV.1.7
B_0	brightness of electron gun	IV.1.8
b_c	coherent scattering length for neutrons	II.5

Symbol	Designation	Chapter
b_h	reciprocal vector	IV.1.4
B_i	relative optimal-frequency dielectric impermeability	I.1/2
$\boldsymbol{b}_i$	base vectors of the reciprocal lattice	IV.1.1
C	atomic concentration	II.2
c	speed of light	I.3, I.6, II.5, I.1/2
$C_0^{(1)}$, $C_0^{(2)}$	Bloch wave amplitudes	IV.1.5
$C(a, \Delta, \boldsymbol{g})$	damping envelope due to beam convergence	IV.1.1.2
$C(t)$	normalized correlation of $S(t)$ and $S_0(t)$	III.1
C_c	chromatic aberration coefficient	IV.1.8, IV.1.4
C_i	polarization coefficient corresponding to polarization state i	II.5
C_S	spherical aberration coefficient	IV.1.7, IV.1.1.2, IV.1.8, IV.1.4
D	coefficient of quasi-elastic force	I.1/2
D	condenser zone plate diameter	II.1
$\boldsymbol{D}$	dielectric displacement	I.1/2
D	dimension of elementary detector	II.3
D	distance from the source to the object	II.4
D	total dose	IV.1.3
d	distance along crystal axis	IV.1.4
d	distance from the object to the recording plane	II.4
$\boldsymbol{d}$	interplanar spacing	IV.1.5
d	one-dimensional periodicity; repeat distance of atoms in a column	IV.1.1.2
d	picture element size	II.2
d	pinhole diameter	II.1
d	step height, film thickness	IV.1.6
d	thickness of surface layer	III.1
d	thickness; distance	I.1/2
$D(a, \Delta, \boldsymbol{g})$	total damping envelope	IV.1.1.2
D_0	dose (photons per pixel dimension)	II.3
d_c	resolution limit	I.1/2
d_g	interplanar spacing of the active reflection	IV.1.1
d_H	interplanar spacing which indices h, k, l	IV.1.1
d_o	distance of object	I.1/2
d_o', d_o''	distance of image	I.1/2
Dz	defocus	IV.1.8
Dz_{gab}	Gabor focus	IV.1.8
Dz_{opt}	optimum holography focus	IV.1.8
Dz_{scherz}	Scherzer focus	IV.1.8
E	accelerating potential	IV.1.1
$\boldsymbol{E}$	electric field strength	I.1/2
E	electron energy	IV.1.6, IV.1.4
E	energy	IV.1.1.2
E	Energy eigenvalue	IV.1.5
E	Young's modulus	III.1
e	electron charge	IV.1.7, IV.1.1, IV.1.1.2, IV.1.8, IV.1.4
e	unit vector normal to a set of planar interfaces	IV.1.1

Symbol	Designation	Chapter
$\boldsymbol{E}_0$	electric field amplitude	I.1/2
$E_{\mathrm{chrom}}(\boldsymbol{R})$	chromatic envelope function	IV.1.8
$E_{\mathrm{coh}}(\boldsymbol{R})$	coherence envelope function	IV.1.8
e_{g}	d-electron orbitals	I.1/2
E_i	energy of column i	IV.1.1.2
E_n	energy of state n	IV.1.1.2
$\boldsymbol{e}_{\mathrm{n}}$	unit normal	IV.1.1
$\boldsymbol{E}_{\mathrm{o}}$	Amplitude of incident beam	II.5
eU	electron energy	IV.1.8
$E_{\mathrm{x,\,y,\,z}}$	components of the electric field amplitude	I.1/2
F	field strength at the cathode	IV.1.6
$\mathscr{F}$	Fourier transform operator	III.1
F	mapping	I.4
f	focal length	II.1, II.4
f	frequency of sound waves	III.1
$f(\boldsymbol{q})$	scattering length	II.5
$F()$	Fourier transform of F(x)	I.4
$\mathscr{F}^{-1}$	inverse Fourier transform operator	III.1
$\mathrm{f}(\theta)$	complex function	IV.1.7
$f(x)$	pupil function	I.4
f, f'	focal distance	I.1/2
f_1, f_2	focal length	I.4
$\mathscr{F}^2$	structure factor	IV.1.6
F_{A}	figure of merit for a pulse	III.1
f_{at}	atomic scattering factor for X-rays	II.5
f_{c}	focal length of construct	II.4
$f_{\mathrm{e}}(\theta)$	atomic scattering factor for electrons	IV.1.1
F_{g}	structure factor of reflection $\boldsymbol{g}$ with indices (h_1, h_2, h_3)	IV.1.1
F_{g}	structure factor of corresponding to a particular reflection	IV.1.5
$F_{\boldsymbol{g}}()$	structure factor	IV.1.1.2
F_{H}	HOLZ reflection structure factor	IV.1.5
$F_{\boldsymbol{h}}$	structure factor for reflection $\boldsymbol{h}$ or $h\ k\ l$	II.5
$f_i(\boldsymbol{g})$	atom scattering factor	IV.1.1.2
f_{Ob}	focal distance of objective	I.1/2
f_{oc}	focal distance of ocular	I.1/2
$\mathscr{F}_{\boldsymbol{R}}$	Fourier transform	IV.1.1.2
$f_X(\theta)$	atomic scattering factor for X-rays	IV.1.1
G	radius of HOLZ ring	IV.1.5
$\boldsymbol{g}$	diffraction vector	IV.1.2
$\boldsymbol{g}$	reciprocal lattice vector	IV.1.1
g	spatial frequency	IV.1.1.2
$\boldsymbol{g}$	spatial frequency vector	IV.1.1.2
$g(t, z)$	distortion of the wave shape due to defocus and attenuation	III.1
$\boldsymbol{g}$, g′	reciprocal lattice vectors	IV.1.5
$\mathscr{G}^2$	lattice factor	IV.1.6
g_{A}	radius of the physical aperture	IV.1.1.2
H	Hamiltonian	IV.1.1.2
$\boldsymbol{H}$	indices h, k, l (or h_1, h_2, h_3)	IV.1.1
H	layer spacing of reciprocal lattice	IV.1.5

Symbol	Designation	Chapter
$\boldsymbol{H}$	magnetic field strength	I.1/2
$\boldsymbol{h}$	diffraction vector or position vector in reciprocal space	IV.1.1
h	Dirac constant	IV.1.1.2
h	distance between the reference and object rays	II.4
h	Planck's constant	II.2, IV.1.4, IV.1.5, IV.1.7, I.6, IV.1.1, IV.1.6
$\boldsymbol{h}$	reciprocal lattice vector corresponding to (h k l) lattice planes	II.5
h	step height	IV.1.2
$\hbar$	Planck constant	IV.1.1.2
h, k, l	Miller indices	
$\boldsymbol{H}_0$	magnetic field amplitude	I.1/2
$\boldsymbol{I}$	intensity	I.1/2, IV.1.3
I	intensity distribution	IV.1.8
I	intensity in image plane	IV.1.1.2
i	imaginary unit	I.4
$I(\boldsymbol{h})$	scattered intensity for scattering vector $\boldsymbol{h}$	IV.1.1
$I(x, y, \Delta E)$	intensity within 3D space defined by x, y and ΔE	IV.1.3
$\boldsymbol{I}_0$	incident intensity	II.3
$\boldsymbol{I}_0$	intensity of beam in air	I.1/2
I_0	intensity of zero-loss peak	IV.1.3
I_0	irradiance (W cm^{-2})	I.3
$I_{00}(E)$	intensity of the incident beam	IV.1.6
$\boldsymbol{I}_1$, $\boldsymbol{I}_2$	intensity of components 1 and 2	I.1/2
I_1^{fl}	intensity of fluorescence emission	I.4
I_1^{refl}	intensity of reflected wave	I.4
I_{coh}	coherent electron current	IV.1.8
I_{g}	diffracted beam intensity	IV.1.5
I_{g}	intensity of the diffracted beam	IV.1.1
I_{i}	image intensity	IV.1.7
$\boldsymbol{I}_{\mathrm{i}}$	intensity of the incident beam	I.1/2
I_{inc}	incident beam current	IV.1.3
$\boldsymbol{I}_{\mathrm{max}}$	maximum intensity	I.1/2
$\boldsymbol{I}_{\mathrm{min}}$	minimum intensity	I.1/2
$\boldsymbol{I}_{\mathrm{mk}}$	intensity (at position m, measured by the detector k)	II.3
I_{o}	incident intensity	II.5
I_{o}	intensity of a direct beam	IV.1.5
$\boldsymbol{I}_{\mathrm{refl}}$	intensity of reflected beam	I.1/2
$\boldsymbol{I}_{\mathrm{res}}$	resulting intensity	I.1/2
I_{S}	intensity of scattered beam	IV.1.1
I_{T}	intensity of transmitted beam	IV.1.1
J	primary electron flux	IV.1.3
J_1	first order Bessel function	I.3, IV.1.1.2
J_1	primary electron flux in the PEELS-STEM	IV.1.3
J_2	primary electron flux in the EFTEM	IV.1.3
j_{coh}	coherent electron current density	IV.1.8
K	amplitude PSF	I.4
$\boldsymbol{K}$	beam direction	IV.1.1.2

Symbol	Designation	Chapter
K	monochromator dispersion	II.4
K	phonon wavenumber	III.1
$\boldsymbol{K}$	phonon wavevector	III.1
K	radius of the Ewald space	IV.1.1
$\boldsymbol{K}$	wavevector of incident electron beam in the crystal corrected for refraction by the mean inner potential	IV.1.1
k	absorption coefficient	I.1/2
k	electron wavelength	IV.1.1.2
k	element detector	II.3
k	wave number	III.1, IV.1.8, IV.1.5
$\boldsymbol{k}$	wave vector	IV.1.1
$\boldsymbol{k}$	wave vector of diffracted wave	II.5
$\boldsymbol{k}$, $\boldsymbol{k}_{res}$, $\boldsymbol{k}_1$, $\boldsymbol{k}_2$	wave vectors	I.1/2
$\boldsymbol{K}_0$	wave vector of indirect electron beam in vacuum	IV.1.1
$\boldsymbol{k_0}$	wave vector for incident wave	II.5
$\boldsymbol{k}_0$	wave vector of indirect beam in crystal	IV.1.1
$\boldsymbol{k}_{0,t}$ $\boldsymbol{K}_{0,t}$	tangential component of $\boldsymbol{k}_0$, $\boldsymbol{K}_0$	IV.1.1
k_1, k_2	incident, scattered photon wavenumber	III.1
$\boldsymbol{k}_1$, $\boldsymbol{k}_2$	incident, scattered photon wavevector	III.1
$\boldsymbol{k}^A$, $\boldsymbol{k}_1^A$, $\boldsymbol{k}_2^A$	wave vectors in the analyzer plane	I.1/2
$\boldsymbol{k}_g$	wave vector of diffracted beam in crystal	IV.1.1
k_R	wave number of Rayleigh waves	III.1
$\boldsymbol{K}_t$	tangential component of $\boldsymbol{K}$	IV.1.1
l	angular momentum quantum number	IV.1.5
l	character of elongation	I.1/2
M	magnification	I.4, II.2, II.3
M	*Mallard*'s constant	I.1/2
M	molar mass	II.1
M	real number in denominator of pseudo inverse filter	III.1
M	target electron mass	IV.1.1.2
m	diffractive order	II.1
m	electron mass	IV.1.1.2
m	mass	I.1/2, II.5, IV.1.5
m	mass of electron	IV.1.7
m	relativistic mass of electron	IV.1.4
$M(s, z)$	transfer matrix of perfect crystal slab	IV.1.1
m_A	fraction of columns of A atoms	IV.1.1.2
m_B	fraction of columns of B atoms	IV.1.1.2
m_e	mass of an electron	I.1/2
M_H	holographic magnification	II.4
M_O	optical magnification	II.4
m_o	rest mass	IV.1.5
M_{obj}	magnification of object	I.1/2
M_{oc}	magnification of ocular	I.1/2
M_t	magnification of tube	I.1/2
N	atomic density	IV.1.4
N	noise	IV.1.1.2

Symbol	Designation	Chapter
N	number of atoms	II.1
N	number of electrons per resolved pixel	IV.1.8
N	number of molecules in the probed volume, V (–)	I.3
N	number of pixels	II.3, IV.1.3
N	number of atoms per unit specimen area	IV.1.3
N	number of unit cells	IV.1.1
N	number of wavelengths	II.4
n	an integer	I.3
n	image order of dislocation	IV.1.1
n	index of refraction	II.4, IV.1.8, I.1/2, I.4
n	number of energy loss channels	IV.1.3
n	number of spatial frequencies	IV.1.1.2
n	principal quantum number	IV.1.5
n	real part of the refractive index; zone number	II.1
$\tilde{n}$	complex refractive index	II.1
N'	number of elctrons statistically significantly recorded per resolved pixel	IV.1.8
n', n''	refractive indices of reference media 1 and 2	I.1/2
N^0	total atomic density	II.2
n_0	number of incident photons	II.3
n_0	refractive index of a reference medium (air)	I.1/2
NA	numerical aperture	I.3, I.4
N_A	Loschmidt constant	II.1
n_α	smallest principal refractive index of the indicatrix	I.1/2
$n_{\alpha'}$	smallest refractive index of an arbitrary section of the indicatrix	I.1/2
n_β	intermediate principal refractive index of the indicatrix, optical normal	I.1/2
$n_{\beta'}$	intermediate refractive index of an arbitrary section of the indicatrix	I.1/2
N_e	number of electrons	I.1/2
n_e	refractive index of extraordinary beam	I.1/2
N_{fringe}	number of hologram fringes	IV.1.8
n_γ	largest principal refractive index of the indicatrix	I.1/2
$n_{\gamma'}$	largest refractive index of an arbitrary section of the indicatrix	I.1/2
n_i	refractive index of medium in which the incident beam propagates	I.1/2
N_{int}	area under rocking-curve	II.5
n_θ	number of observed isochromates	I.1/2
n_k	komplex units	I.1/2
n_o	refractive index of ordinary beam	I.1/2
N_{pix}	number of pixels for hologram detection along one direction of detector	IV.1.8
n_{refr}	refractive index of refracting medium	I.1/2
$o(x, y)$	object wave	IV.1.8
$o_r(x, y)$	reconstructed object wave	IV.1.8
$\boldsymbol{P}$	dielectric polarization tensor	I.1/2
P	Poisson distribution	IV.1.3
$P(\theta)$	pupil function of the lens	III.1
$\boldsymbol{q}$	reciprocal vector	IV.1.1.2
$\boldsymbol{q}$	scattering vector	II.5
$\boldsymbol{q}$	superlattice satellite reflection	IV.1.5
q	tube factor	I.1/2
$Q(t)$	substitution function in Fourier expression for V	III.1

Symbol	Designation	Chapter
$QQ(t)$	autocorrelation function of $Q(t)$	III.1
R	displacement field	IV.1.5
$\boldsymbol{R}$	displacement vector	IV.1.2
R	Parameter for modelling background	IV.1.3
R	reflectivity	I.1/2
$\boldsymbol{R}$	spatial frequency vector	IV.1.8
$\boldsymbol{R}$	vector in the plane perpendicular to z	IV.1.1.2
r	dimension of pixel	II.3
r	pixel size (resolution) (μm)	II.3
r	point in space	II.5
r	position vector	IV.1.5
r	radius	IV.1.1.2
r	reference wave	IV.1.8
r	resolution	II.4
r	resolving power (t/N)	II.3
$\boldsymbol{R}_{\parallel,\perp}$	radial, normal component of $\boldsymbol{R}(\boldsymbol{r})$	IV.1.1
R', R''	reflectivities of sample in contact with reference media of n' and n''	I.1/2
r'_h	integrated reflectivity for extended crystal	II.5
$R(\theta)$	reflectance function of the specimen	III.1
$\boldsymbol{R}(\boldsymbol{r})$	displacement field of defect	IV.1.1
R_0	radius of laser beam	I.3
r_0	classical electron radius	II.1
r_A	aperture radius	IV.1.6
R_B	Bohr radius	IV.1.1.2
R_C	carrier frequency	IV.1.8
r_c	classical electron radius	II.5
r_d	resolution limit	II.2
r_g	lateral resolution	II.3
r_h	integrated reflectivity for sample entirely bathed in incident beam	II.5
$\boldsymbol{R}_i$	position of the column i	IV.1.1.2
r_L	longitudinal resolution	II.4
$\boldsymbol{r}_{\underline{L}}$	lattice vector	IV.1.1
R_{lim}	information limit	IV.1.8
R_{max}	maximum spatial frequency of image wave	IV.1.8
R_{min}	lower limit of Scherzer band	IV.1.8
r_n	radius of the nth zone	II.1
$\boldsymbol{r}_n$	position of the n-th scattering vector	IV.1.1
R_σ	polynomial function of Poisson ratio	III.1
R_{scherz}	upper limit of Scherzer band (point resolution)	IV.1.8
r_T	transverse resolution	II.4
R_x, R_y	components of $\boldsymbol{R}$ in directions x, y	IV.1.8
S	amplitude of scattered beam	IV.1.1
S	signal	IV.1.1.2
S	signal intensity	I.3
S	string's strength parameter	IV.1.5
s	electron path	IV.1.8
s	path length	II.3
s	size of source	II.3, II.2
S^-	amplitude of scattered beam with $s \Rightarrow -s$	IV.1.1

Symbol	Designation	Chapter
$\bar{s}(f), \bar{S}(f), \bar{S}_0(f)$	Fourier transform of the corresponding functions in the time domain	III.1
$\bar{S}_2(f)$	filtered signal	III.1
S', S'', T', T''	S, T in the two-beam approximation	IV.1.1
$S(t)$	specimen signal	III.1
$s(t)$	two-way lens response	III.1
$S_0(t)$	reference signal	III.1
s_a, s_b, s_c	size of turning points of a pulse	III.1
S_g	modulus of the amplitude of ϕ_g	IV.1.1
s_g	deviation parameter for reflection g	IV.1.5, IV.1.1
$s_{\boldsymbol{g}}$	excitation error	IV.1.1.2
$s_{g,e}$, s_{eff}	effective local s value in a deformed crystal	IV.1.1
s_λ	detector sensitivity at wavelength λ	I.3
$\mathscr{T}$	amplitude transmission	II.1
T	amplitude of transmitted beam	IV.1.1
T	total recording time	IV.1.3
t	distance travelled by electron	IV.1.7
t	sample or specimen thickness	II.5, II.2, IV.1.5, II.1, IV.1.1.2
t	time	I.6, II.3, IV.1.8, IV.1.1.2, I.1/2
t	typical dimension of region to be constructed	II.3
T^-	amplitude of transmitted beam with $s \Rightarrow -s$	IV.1.1
$T()$	transfer function	IV.1.1.2
$t(\boldsymbol{R})$	Fourier transform of the transfer function	IV.1.1.2
$t/t_{0.1}$	thickness ratio of specimen	IV.1.4
t_1, t_2, t_3	reflection times	III.1
t_{2g}	d-electron orbitals	I.1/2
t_a, t_b, t_c	time of turning points of a pulse	III.1
T_c	critical temperature	IV.1.1.2
t_g, t_{-g}	extinction distance of reflection $\boldsymbol{g}$, $-\boldsymbol{g}$	IV.1.1
T_λ	instrument throughput at wavelength λ	I.3
t_m	time at which the pulse maximum occurs	III.1
U	accelerating voltage	IV.1.8
$\boldsymbol{u}$	displacement	II.5
u	normalized distance	I.3
$U()$	electrostatic potential	IV.1.1.2
u, t	substitution variables in Fourier expression for V	III.1
U, U_0, U_1	disturbance of the electric field	I.4
u_1	range of kz in $V(z)$ data	III.1
$u_\nu(T)$	spectral energy density for a blackbody	I.6
$U^n(\boldsymbol{R})$	two-dimensional conditional potential	IV.1.5
V	modulus of coherence; contrast of hologram fringes	IV.1.8
V	video signal	III.1
v	normalized distance to optical axis	I.3
v	sample volume	II.5
v	velocity	IV.1.1.2
$V(f)$	video signal as an explicit function of frequency	III.1
$V(\boldsymbol{R})$	projected potential	IV.1.1.2
$V(\boldsymbol{r})$	electrostatic potential field	IV.1.8

Symbol	Designation	Chapter
$V(\boldsymbol{r})$	lattice potential	IV.1.1
$V(\boldsymbol{R}, z)$	crystal potential	IV.1.1.2
$V(z)$	video signal as an explicit function of defocus	III.1
$V(z, s)$	vacuum matrix	IV.1.1
V^*	volume of the reciprocal unit cell	IV.1.1
v, $\boldsymbol{v}$	velocity of electron	IV.1.7
V_0	constant part of the lattice potential: mean inner potential	IV.1.1
V_0	inner potential	IV.1.6
v_0	velocity of sound waves in water	III.1
V_C	unit cell volume	IV.1.1, II.5
$V_{def,\ deformed}$	potential of a deformed lattice	IV.1.1
V_{el}	electric Potential	IV.1.8
V_g	Fourier coefficient of lattice potential	IV.1.1
$V_{\boldsymbol{g}}$	structure factor	IV.1.1.2
v_g	Fourier coefficient of crystal potential	IV.1.5
V_G, V_R	contribution of geometrical ray, Rayleigh ray to the video signal (including phases)	III.1
V_h	Fourier coefficient of crystal potential	IV.1.4
V_{inel}	contrast reduction of hologram fringes by inelastic scattering	IV.1.8
V_{inst}	contrast reduction of hologram fringes by instabilities	IV.1.8
V_L	video function with lead as specimen	III.1
v_l	velocity in a cell or surface layer	III.1
V_{MTF}	contrast reduction of hologram fringes by MTF of detector	IV.1.8
V_o	mean crystal potential	IV.1.5
V_p	projected potential	IV.1.8
$V_p(\boldsymbol{R})$	projected potential of an object	IV.1.1.2
v_R	Rayleigh velocity	III.1
v_s	shear velocity	III.1
w	width of hologram	IV.1.8
$W(\boldsymbol{r})$	imaginary part of the lattice potential	IV.1.1
W_{420}, δW_{400}	parameters to describe texture in pole figures of X-ray or neutron activation measurements	III.1
w_g	deviation parameter ($= s_g \xi_g$)	IV.1.5
x, y	coordinates of pixel position	IV.1.3
X, Y, Z	principal axes of the indicatrix	I.1/2
x, y, z	coordinates	IV.1.7, I.1/2
x_0, x_1	coordinates	I.4
x_i	principal axes of a coordinate system ($i = 1, 2, 3$)	I.1/2
x_{img}, x_{obj}	lateral distance in the image/object	I.4
x_k, y_k, z_k	komplex units	I.1/2
y_0, y_1	coordinates	I.4
Z	atomic number	II.5, IV.1.1.2
Z	mean atomic number	IV.1.5
z	axial altitude of sample slice	I.3
z	depth in crystal	IV.1.5
z	distance along z axis (cm)	I.3
z	distance between specemen and focal plane	III.1
z_0, t	total crystal thickness	IV.1.1
Z_0, Z_s	impedances of coupling fluid and substrate	III.1

Symbol	Designation	Chapter
Z_1	impedance of a cell or surface layer	III.1
Z_2	atomic number	IV.1.4
z_{img}, z_{obj}	axial distance in the image/object	I.4
$(\boldsymbol{R})$	eigenstate	
a	angular aperture	I.3
a	aperture angle	IV.1.4
a	beam convergence angle	IV.1.1.2
a	decay of Rayleigh oscillations	III.1
a	diffraction angle	II.4
a	half the opening angle of objective lens	I.4
a	incident beam convergence angular	IV.1.5
a	smallest refractive index of the indicatrix	I.1/2
a_0	attenuation of sound waves in water	III.1
a_a, a_p	axial and planar channeling parameters	IV.1.4
a_g	phase difference introduced by translation interfaces	IV.1.1
a_i	base vectors of the crystal lattice	IV.1.1
a_{img}, a_{obj}	viewing angle of the image/object	I.4
a_l	attenuation in a cell or surface layer	III.1
a_N	normalized attenuation	III.1
a_R	attenuation of Rayleigh waves	III.1
β	angle of deflection	IV.1.7
β	bandwidth of the transducer	III.1
β	intermediate refractive index of the indicatrix	I.1/2
β	numerical coefficient	II.4
β	scattering angle	IV.1.3
χ	index of absorption	I.1/2
χ	wave aberration	IV.1.8
$\chi(\boldsymbol{g})$	phase shift	IV.1.1.2
$\chi(\boldsymbol{R})$	wave aberration	IV.1.8
$\chi(x, y, z)$	phase shift	IV.1.1.2
χ_a	antisymmetric wave aberration	IV.1.8
χ_{mic}	wave aberration in microscope	IV.1.8
χ_{num}	wave aberration generated in computer	IV.1.8
χ_s	symmetric wave aberration	IV.1.8
Δ	diffraction limit	II.1
Δ	energy window	IV.1.3
Δ	Laplacean operator; half-width	IV.1.1.2
Δ	measure of defocus	IV.1.2
Δ	moiré wavelength	IV.1.1
Δ	path difference	II.4
Δ	Pendellösung period	II.5
Δ	principal double refraction, birefringence	I.1/2
Δ	size of diffraction-limited focused probe	IV.1.5
δ	defocus	IV.1.7
δ	efficiency influenced by diffraction (–)	I.6
δ	excitation difference at domain boundary	IV.1.1
δ	resolution	IV.1.6
Δ'	partial birefringence of arbitrary indicatrix section	I.1/2
$\delta(\boldsymbol{g})$	Dirac delta function	IV.1.1.2

Symbol	Designation	Chapter
Δa	change in the total attenuation	III.1
δ_C	chromatic resolution	IV.1.6
δc	accuracy of wave aberration	IV.1.8
δ_{chr}, δ_{diff}, δ_{sph}	chromatic, diffraction, spherical aberrations	IV.1.4
δC_s	accuracy of spherical aberration	IV.1.8
δ_D	diffraction disk resolution	IV.1.6
δ_d	diffraction limit	II.2
δDz	accuracy of defocus	IV.1.8
Δ^E	energy loss	IV.1.3
ΔE	electron energy and its fluctuation	IV.1.4
ΔE	energy range	IV.1.8
Δf	defocusing distance	IV.1.5
Δf	focus	IV.1.1.2
Δf	period of oscillation	III.1
δf	phase detection limit of reconstructed wave	IV.1.8
$\Delta \boldsymbol{g}_{\parallel, \perp}$	components of $\Delta \boldsymbol{g}$	IV.1.1
Δ_J	particular birefringence	I.1/2
Δn	resolution	I.6
Δ_p	parallel moiré wavelength	IV.1.1
δ_p	width due to penumbral blurring	II.2
Δq	angular deviation from the Bragg conclition unstrained angle of inclination of the diffractions plan	IV.1.5
Δq	angular step	II.3
Δq	broadening of $R(t)$	III.1
$\Delta_{\boldsymbol{R}}$	Laplacean operator acting in the plane $\boldsymbol{R}$	IV.1.1.2
Δ_r	rotation moiré wavelength	IV.1.1
δ_S	spheral resolution	IV.1.6
δ_T	total aberration	IV.1.4
Δt_1	recording time pen pixel in the PEELS-STEM	IV.1.3
Δx	lateral point resolution	I.4
Δz	depth of focus	I.1/2
Δz	oscillation period of $V(z)$	III.1
Δz_{fl}, Δz_{ref}	axial point resolution for fluorescence/reflection	I.4
ε	Darwin width	II.5
ε	defocus	IV.1.1.2
ε	dielectric permittivity	I.1/2
ε	ellipticity of illumination	IV.1.8
ε_0	specific dielectric permittivity	I.1/2
$\varepsilon_{i,j}$	principal constants of ε	I.1/2
ε_s	Scherzer defocus	IV.1.1.2
Φ	phase of image wave	IV.1.8
ϕ	angle	I.1/2
ϕ	phase of object wave	IV.1.8
ϕ (X)	phase difference	IV.1.7
$\phi(\boldsymbol{R})$	image wavefunction	IV.1.1.2
$\phi(x)$	phase angle	III.1
$\phi^*(\boldsymbol{g})$	complex conjugate of wavefunction in Fourier space	IV.1.1.2
ϕ_g	amplitude of ψ_g	IV.1.1
ϕ_i	incident angle	I.1/2

Symbol	Designation	Chapter
ϕ_i	localized wavefunction	IV.1.1.2
$\phi_i(\boldsymbol{R})$	wavefunction of column i	IV.1.1.2
ϕ_{ic}	critical angle of incidence (*Brewster*'s angle)	I.1/2
Φ_m	magnetic flux	IV.1.8
$\phi_n(\boldsymbol{R})$	n-th eigenstate	IV.1.1.2
ϕ_{refl}	angle of reflection	I.1/2
ϕ_{refr}	angle of refraction	I.1/2
ϕ_{tot}	angle of total reflection	I.1/2
Γ	path difference	I.1/2
γ	largest refractive index of the indicatrix	I.1/2
γ	relativistic mass factor	IV.1.5
γ	stacking fault energy	IV.1.1
γ, d	empirical parameters in the figure of merit F_A	III.1
Γ_0	magnification factor	I.3
Γ_θ	aperture-angle dependent path difference	I.1/2
η	draw ratio of an alloy	III.1
η	natural mismatch	IV.1.5
η	phase shift coefficient	II.1
θ	aperture angle	I.1/2
θ, ϕ	pair of angles in the backfocal plane	I.4
φ_g	diffracted beam amplitude	IV.1.5
λ	wavelength	IV.1.4, IV.1.5, IV.1.6, I.1/2, I.4, I.5, II.1, II.4, II.5, IV.1.1.2, IV.1.2, I.3, I.6, IV.1.7, IV.1.1, II.2
$\lambda'(x, y, z)$	wavelength at position x y z	IV.1.1.2
λ, λ_0	electron wavelength (in vacuum)	IV.1.8
λ_0	wavelength of monochromatic light	I.1/2
λ_0	wavelength of sound waves in water	III.1
λ_c	wavelength of construct	II.4
λ_f	electron wavelength in matter	IV.1.6
λ_i	wavelength of incident beam	I.1/2
λ_i	inelastic mean free path	IV.1.3
λ_R	wavelength of Rayleigh waves	III.1
λ_{refr}	wavelength of refracted beam	I.1/2
μ	absorption coefficient	II.2, II.3
μ	complex degree of coherence	IV.1.8
μ	magnetic permeability	I.1/2
μ	mass absorption coefficient	II.1
μ	shear modulus	III.1, IV.1.1
$\mu(\boldsymbol{R})$	absorption function	IV.1.1.2
$\mu(x, y)$	absorption coefficient	IV.1.1
μ_0	specific magnetic permeability	I.1/2
μ_1	linear absorption coefficient	II.1
μ_a	index of absorption	I.1/2

Symbol	Designation	Chapter
ν	frequency	I.1/2, IV.1.6, I.6
ν	Poisson's ratio	IV.1.1, IV.1.5
ν	spatial frequency	II.4
$\bar{\nu}$	wave number (cm^{-1})	I.6
ν_0	own resonance frequency	I.1/2
ν_p	plasma frequency	I.1/2
θ	angle of incidence	III.1
θ	angle of incidence in grating's law	II.4
θ	angular variable	IV.1.7
θ	Bragg angle	II.5
θ	diffraction angle	IV.1.1.2
θ	limiting optical throughput	I.6
θ_0	limit of validity of the function $R(t)$	III.1
θ_B	Bragg angle	IV.1.2
θ_C	illumination aperture	IV.1.8
θ_c	illumination aperture (half angle)	IV.1.8
θ_g	phase angle of Fourier coefficient V_g	IV.1.1
θ_H	Bragg angle	IV.1.1
θ_n	Bragg angle of order n	IV.1.1
θ_R	Rayleigh angle	III.1
θ_s	shear critical angle	III.1
ρ	density; reflectivity	II.1
ρ	distance between two object points	II.4
ρ	resolution	IV.1.1.2
$\rho\,(\boldsymbol{q})$	Fourier transform of the electron density	II.5
$\rho_{\boldsymbol{h}}$	Fourier component with wave vector $\boldsymbol{h}$ of electron density in crystal	II.5
ρ_l	information limit	IV.1.1.2
$\rho_p(\boldsymbol{R})$	projected charge density	IV.1.1.2
ρ_s	point resolution	IV.1.1.2
Σ	summation symbol	IV.1.1.2
σ	attenuation cross section	II.2
σ	cross-section	IV.1.3
σ	half angle of most oblique light rays	I.1/2
σ	interaction parameter	IV.1.8
σ	Poisson ratio	III.1
$\sigma(\alpha)$	shift matrix	IV.1.1
σ_a	atomic cross section for photoelectric absorption	II.1
σ_g	reduced dynamical deviation parameter of reflection $\boldsymbol{g}$	IV.1.1
$\sigma_{g,i}$ $\sigma_{g,r}$	imaginary, real part of σ_g	IV.1.1
σ_i	imaginary part of σ_g	IV.1.1
σ_i	Flynn occupation parameter	IV.1.1.2
σ_l	differential cross section (cm^2 $sterad^{-1}$ $molecule^{-1}$)	I.3
σ_M	matrix cross section	II.2
σ_r	real part of σ_g	IV.1.1
σ_{xx}, σ_{yy}	surface stresses in the x- and y-directions	III.1
τ	counting time	IV.1.3
τ	transmission	I.1/2
τ, $t(\boldsymbol{g})$	transmission cross-coefficient	IV.1.1.2

Symbol	Designation	Chapter
τ_g	absorption length of reflection $\boldsymbol{g}$	IV.1.1
υ	frequency	II.2
v	phase velocity, propagation velocity	I.1/2
v_0	velocity of ordinary beam	I.1/2
υ_0, υ_i	absolute wave numbers (cm^{-1})	I.3
v_e	velocity of extraordinary beam	I.1/2
v_i	velocity of propagation rays ($i = 1, 2, 3$)	I.1/2
v_{refr}	velocity of refracted beam	I.1/2
ς_0, ς_l	density of fluid, solid	III.1
Ω	solid angle (sterad)	I.3
$\boldsymbol{\Omega}$	phonon angular frequency	III.1
ω	crystal rotation angle	II.5
ω	noise amplification	II.3
ω	spatial frequency	I.4
$\dot{\omega}$	rocking speed	II.5
ω_1, ω_2	incident, scattered light angle; incident, scattered photon angular frequency	III.1
ξ	conjugate of z	IV.1.1.2
ξ	overall system efficiency (–)	I.6
ξ, x_0	independent variable in Fourier transform of $V(z)$, the peak in the Fourier transform occurs at $\xi = \xi_0$	III.1
ξ_g^{eff}	effective extinction distance	IV.1.5
ξ_g	extinction length	IV.1.5
ξ_H	HOLZ extinction length	IV.1.5
$\boldsymbol{\Psi}$	total electron wave function	IV.1.1
ψ	electron wave function	IV.1.1
ψ	object wavefunction	IV.1.1.2
$\boldsymbol{\psi}(r)$	electron wavefunction	IV.1.5
ψ_0	wave function of electrons in incident beam	IV.1.1
$\boldsymbol{\Psi}_1$, $\boldsymbol{\Psi}_2$	Bloch wave fields in crystal	IV.1.1
ψ_g	wave function of electrons in diffracted beam	IV.1.1
$\psi_g(z)$	amplitude of the beam $\boldsymbol{g}$ at a depth z	IV.1.1.2
$\psi_{S,T}$	ψ referring to plane waves with the amplitudes S, T	IV.1.1
ζ	range of z in measurement of $V(z)$	III.1

Abbreviation	Explanation
1D, 2D, 3D	one-, two-, three-dimensional
3D	three dimensional
ANN	artificial neural network
ATR	attenuated total reflectance
B.F.	bright field
BZ	Brillouin zone (boundary)
CCD	charge-coupled device
CORALIS	confocal Raman line scanning
CPEM	contact photoelectron X-ray microscope
CVD	chemical vapor deposition
D.F.	dark field

Abbreviation	Explanation
dB	decibel, unit of attenuation expressed as 10 times the logarithm to base 10 of the ratio of powers
DFI	direct Fourier inversion
DQE	detection quantum efficiency
EDX	energy-dispersive X-rays
EII	electron iradiation induced
ELNES	energy loss near edge structure
EM	electron microscope
EVA	ethylene vinyl acetate copolymer
EXELFS	extended electron energy-loss fine structure
f.c.c	face-centered cubic
FBP	filler back projection
FEG	field emission gun
FET	field-effect transistor
FITC	fluorescein isothiocyanate
FOLZ	first order Laue zone
FS	final start
FT	Fourier transform
FTIR	Fourier transform infrared
FWHM	full width half maximum
FZP	Fresnel zone plate
G. P.	Guinier-Preston
GIF	Gatan imaging filter
h.c.p.	hexagonal close-packed
HOLZ	higher-order Lane zone
HREM	high-resolution electron microscopy
IR	infrared
IRF	impulse response function
KDI	Kirchhoff diffraction integral
LCD	liquid crystal display
LDPE	low density polyethylene
LEED	low-energy electron diffraction
LHS	left hand side
LQ	liquid quenching
MCT	nitrogen cooled mercury cadmium-telluride photodetector
MEM	maximum entropy method
MLS	multiple least-squares
Mrayl	10^6 kg m^{-2} s^{-1} (unit of acoustic impedance)
MTF	modulation transfer function
NA	numerical Aperture
NEP	noise-equivalent power of the detector
NEXAFS	near-edge X-ray absorption fire structure
Np	neper, unit of attenuation expressed as the logarithm to base e of the ratio of amplitudes
NRVO	nonresonant vertical osscilation
PCD	projected charge density
PET	poly(ethylene terephtalate)
PMMA	polymethylmethacrylate
PSF	point spread function
PTFE	polytetrafluoroethylene (Teflon)

Abbreviation	Explanation
PTZ	lead titanate-zirconate ($PbZrO_3$-$PbTiO_3$)
RH	right-handed
RHEED	reflection high energy electron diffraction
RHS	right hand side
SAED	selected area electron diffraction
SAFT	synthetic aperture Fourier transform
SEI	secondary electron imaging
SFT	stacking fault tetrahedron
SNR	signal-to-noise ratio
SR	synchroton radiation
SR	synchrotron radiation
STEM	scanning transmission electron microscope
STOM	scanning tunneling optical microscopes
TCC	transmission cross-coefficient
TEM	transmission electron microscopy
UFM	ultrasonic force microscope
UHV	ultrahigh vacuum
UV	ultraviolet
VLSI	very low scale integrated circuits
VUV	vacuum ultraviolet
WPO	weak phase object
WTF	wave transfer function
ZOLZ	Zero order Laue zone

Techniques	Explanation
AEEM	Auger electron emission microscopy
AEM	analytical electron microscopy
AES	atomic emission spectroscopy
AES	Auger electron spectroscopy
AFM	atomic force microscopy
APFIM	atom probe field ion microscopy
ARPES	angle-resolved photoelectron spectroscopy
ATRS	attenuated total reflectance spectroscopy
BEEM	ballistic electron emission microscopy
BEES	ballistic electron emission spectroscopy
BF CTEM	bright field conventional transmission electron microscopy
CBED	convergent beam electron diffraction
CBRHEED	continuous beam reflective high-energy electron diffraction
CCSEM	computer-controlled scanning electron microscopy
CITS	current imaging tunneling spectroscopy
CL	cathodoluminescence
CLSM	confocal laser scanning microscopy
CT	computer-aided tomography
CTEM	conventional transmission electron microscopy
CVD	chemical vapor deposition
DLTS	deep level transient spectroscopy

Techniques	Explanation
EBT	electron beam testing
EDS	electron diffraction spectrometry
EDS	energy-dispersive spectroscopy
EDX	energy dispersive X-ray spectroscopy
EELS	electron energy loss spectroscopy
EFTEM	energy filtered transmission electron microscopy
EM	electron microscopy
EPMA	electron probe microanalysis
EPXMA	electron probe X-ray microanalysis
ESCA	electron spectroscopy for chemical analysis
ESEM	environmental scanning electron microscopy
ESI	electron spectroscopic imaging
ESI	element-specific imaging
FFM	friction force microscopy
FIB	focused ion beam milling
FIM	field ion microscopy
FMT	fluorescent microthermography
FT-IR	Fourier transform infrared spectroscopy
HREM	high resolution electron microscopy
HRSEM	high resolution scanning electron microscopy
HRTEM	high resolution transmission electron microscopy
HVEM	high voltage electron microscopy
LACBED	large angle convergent beam electron diffraction
LCT	liquid crystal thermography
LEEM	low-energy electron microscopy
LFM	lateral force microscopy
LM	light microscopy
LMMS	laser microprobe mass spectrometry
LOM	light optical microscopy
LPCVD	low-pressure chemical vapor deposition
LTSLEM	low-temperature scanning laser electron microscopy
M-PIXE	micro-(proton-induced X-ray emission spectroscopy)
MBE	molecular beam epitaxy
MEM	mirror electron microscopy
MFM	magnetic force microscopy
MOVPE	metal-organic vapor phase epitaxy
MRI	magnetic resonance imaging
MULSAM	multispectral Auger microscopy
NMR	nuclear magnetic resonance
OM	optical microscopy
PCA	principal components analysis
PEELS	photoelectron energy loss spectroscopy
PEEM	photoemission electron microscopy
PFA	principal factor analysis
PIXE	proton induced X-ray emission spectroscopy
PL	photoluminescence
PPM	point-projection microscopy
RBS	Rutherford backscattering spectroscopy
RDE	reactive deposition epitaxy

Techniques	Explanation
REM	reflection energy microscopy
REM	reflection electron microscopy
RHEED	reflection high-energy electron diffraction
SAM	scanning acoustic microscopy
SAM	scanning Auger microscopy
SAXS	small-angle X-ray scattering
SCM	scanning capacitance microscopy
SDLTS	scanning deep level tansient spectroscopy
SECM	scanning electrochemical microscopy
SEELS	serial electron energy-loss spectroscopy
SEEM	secondary electron emission spectroscopy
SEM	scanning electron microscopy
SEMPA	scanning electron microscopy with polarization analysis
SFM	scanning force microscopy
SIMS	secondary ion mass spectrometry
SLEEM	slow low-energy electron microscopy
SNOM	scanning near-field optical microscopy
SNPM	scanning near-field probe microscopy
SPE	solid phase epitaxy
SPLEED	spin-polarized low-energy electron diffraction
SPLEEM	spin-polarized low energy electron microscopy
SPM	scanning probe microscopy
SPSTM	spin-polarized scanning tunneling microscopy
SQUID	superconducting quantum interference device
SREM	scanning reflection electron microscopy
STEM	scanning transmission electron microscopy
STM	scanning tunneling microscopy
STOM	scanning tunneling optical microscopy
STS	scanning tunneling spectroscopy
STXM	scanning transmission X-ray microscopy
TED	transmission electron diffraction
TEEM	thermionic electron emission microscopy
TEM	transmission electron microscopy
TL	thermoluminescence
TS	tunneling spectroscopy
TSMFM	tunneling stabilized magnetic force microscopy
TXM	transmission X-ray microscopy
UFM	ultrasonic force microscopy
UMT	ultra microtomography
VPE	vapor phase epitaxy
WDS	wavelength dispersive spectroscopy
XES	X-ray emission spectroscopy
XPS	X-ray photoelectron spectroscopy
XPS	X-ray photoemission spectroscopy
XTEM	cross-sectional transmission electron microscopy

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

Index

S. Amelinckx, D. van Dyck, J. van Landuyt,
G. van Tendeloo

Handbook of Microscopy

Methods II

Handbook of Microscopy

Applications in Materials Science, Solid-State Physics and Chemistry

Methods I
1997. ISBN 3-527-29280-2.

Methods II
1997. ISBN 3-527-29473-2.

Applications
1997. ISBN 3-527-29293-4.

Further Reading from VCH

S. N. Magonor, M.-U. Whangbo
Surface Analysis with STM and AFM
Experimental and Theoretical Aspects of Image Analysis
ISBN 3-527-29313-2

D. A. Bonnell
Scanning Tunnelling Microscopy and Spectroscopy
Theory, Techniques and Applications
ISBN 3-527-27920-2

Distribution:
VCH, P.O. Box 10 11 61, D-69451 Weinheim (Federal Republic of Germany)
Switzerland: VCH, P.O. Box, CH-4020 Basel (Switzerland)
United Kingdom and Ireland: VCH (UK) Ltd., 8 Wellington Court, Cambridge CB1 1HZ (England)
USA and Canada: VCH, 333 7th Avenue, New York, NY 10001 (USA)
Japan: VCH, Eikow Building, 10-9 Hongo 1-chome, Bunkyo-ku, Tokyo 113 (Japan)

ISBN 3-527-29473-2

Handbook of Microscopy

Applications in Materials Science, Solid-State Physics and Chemistry

Edited by
S. Amelinckx, D. van Dyck,
J. van Landuyt, G. van Tendeloo

Methods II

Weinheim · New York · Basel · Cambridge · Tokyo

Prof. S. Amelinckx
Electron Microscopy for
Materials Science (EMAT)
University of Antwerp - RUCA
Groenenborgerlaan 171
2020 Antwerp
Belgium

Prof. D. van Dyck
Electron Microscopy for
Materials Science (EMAT)
University of Antwerp - RUCA
Groenenborgerlaan 171
2020 Antwerp
Belgium

Prof. J. van Landuyt
Electron Microscopy for
Materials Science (EMAT)
University of Antwerp - RUCA
Groenenborgerlaan 171
2020 Antwerp
Belgium

Prof. G. van Tendeloo
Electron Microscopy for
Materials Science (EMAT)
University of Antwerp - RUCA
Groenenborgerlaan 171
2020 Antwerp
Belgium

Published by
VCH Verlagsgesellschaft mbH, Weinheim (Federal Republic of Germany)

Editorial Directors: Dr. Peter Gregory, Dr. Ute Anton, Dr. Jörn Ritterbusch
Production Manager: Dipl.-Wirt.-Ing. (FH) Hans-Jochen Schmitt

Library of Congress Card No. applied for.

A catalogue record for this book is available from the British Library.

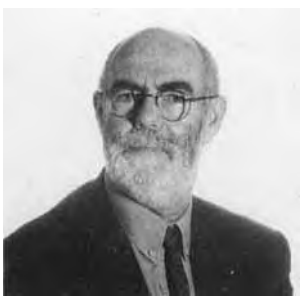
Die Deutsche Bibliothek – CIP-Einheitsaufnahme
Handbook of microscopy : applications in materials science,
solid state physics and chemistry / ed. by S. Amelinckx ... -
Weinheim ; New York ; Basel ; Cambridge ; Tokyo : VCH.
NE: Amelinckx, Severin [Hrsg.]
Methods 2 (1997)
ISBN 3-527-29473-2

Composition: Alden Bookset, England
Printing: betz-druck, D-64291 Darmstadt
Bookbinding: W. Osswald, D-67433 Neustadt

Short biography of the editors

Severin Amelinckx was born in Willebroek, Belgium, in 1922. He studied at the University of Ghent, receiving his first degree (licence) in mathematics in 1944, his doctorate in physics in 1952, and his aggregation in physics in 1955. Currently he is Emeritus Professor of General Physics and Crystallography associated with the EMAT laboratory of the University of Antwerp (RUCA). Until 1987 he was Director General of the Belgian Nuclear Research Establishment at Mol. He is a member of the European Academy and of the Koninklijke Academie voor Wetenschappen, Letteren en Schone Kunsten van België and former chairman of the division of sciences of this academy.

His research interests include electron diffraction contrast imaging, defects in solids, phase transformations and their resulting domain structures, crystal growth, dislocations, fullerenes and nanotubes, the structure of high-T_c superconductors, modulated structures, and order–disorder in alloys.



Joseph Van Landuyt, who was born in St. Amandsberg, Belgium, in 1938, obtained both his licence (1960) and doctorate in physics (1965) from the University of Ghent. At present he is Professor of General Physics and Crystallography at the University of Antwerp (RUCA and UIA) and of Electron Microscopy at UIA and the Flemish University of Brussels (VUB). He is a member of the Koninklijke Academic voor Wetenschappen, Letteren en Schone Kunsten van België.

His research interests are centered on the study of nanostructural features in alloys, ceramics, and minerals (in particular gems), with special interest in defects in semiconductors and their relation to device performance. More general subjects of interest are structural variants, defects, and phase transitions in various solids.

Gustaaf Van Tendeloo, born in Lier, Belgium, in 1950, received his licence in physics from the University of Brussels (VUB) in 1972, his doctorate from the University of Antwerp (UIA) in 1974, and his aggregation from the University of Brussels (VUB) in 1981. He has been associated with the University of Antwerp (RUCA) since 1972, but has spent extended periods of time as a researcher in Berkeley (USA), Caen (France), and elsewhere. He is currently Professor of Solid-State Physics at the University of Brussels (VUB) and of the Physics of Materials at the University of Antwerp (RUCA and UIA).

His research interests include the electron microscopy of inorganic solids (in particular high-T_c superconductors), solid-state phase transitions, modulated structures, fullerenes, defects in crystals order–disorder in alloys, and nanostructural features in general.

Dirk Van Dyck was born in Wilrijk, Belgium, in 1948. He studied physics, receiving his licence from the University of Brussels (VUB) in 1971 before moving to the University of Antwerp (UIA) for his doctorate (1977) and aggregation (1987). He has been associated with the University of Antwerp since 1971, and is at present Professor of Theoretical Mechanics, Digital Techniques and Image Processing at the University of Antwerp.

Among his research interests are theoretical aspects of dynamic electron diffraction and imaging, holographic reconstruction and structural retrieval, image processing and pattern recognition, and artificial intelligence. In particular, he is involved in the development of a 1 Å resolution microscope in the framework of the Brite/Euram program of the European Union.

The four editors belong to the Electron Microscopy for Materials Science (EMAT) laboratory, University of Antwerp (RUCA), which was founded in 1965. All four have contributed significantly to the development of electron microscopy and its application by numerous publications in journals and books and are members of editorial boards of several international journals in the field of materials science. They have also delivered numerous invited lectures at international conferences in their respective areas of research.

List of Contributors

Bauer, Ernst (**V**:3)
Physikal. Institut
Technische Universität Clausthal
38678 Clausthal
Germany

Bonnet, Noél (**VIII**:2)
INSERM Unit 314
University of Reims
21, rue Clément Ader
51100 Reims
France

Cerezo, Alfred; Smith, George D. W. (**VI**:2)
Department of Materials
University of Oxford
Parks Road
Oxford OX1 3PH
United Kingdom

Cory, David G.; Choi, Sungmin (**V**:1)
Dept. of Nuclear Engineering,
NW 14-4111
Massachusetts Institute of Technology
Cambridge, MA 02139
USA

Cowley, John M. (**IV**:2.2)
Arizona State University
Dept. of Physics & Astronomy
Box 871504
Tempe, AZ 8528-1504
USA

DiNardo, N. John (**VII**:4)
Department of Physics and
Atmospheric Science
Drexel University
Philadelphia, PA 19104
USA

Fiermans, Lucien; De Gryse, Roger (**IV**:2.4)
Department of Solid State Sciences
Surface Physics Division
University of Gent
Krijgslaan 281/S1
9000 Gent
Belgium

Herrmann, Karl Heinz (**VIII**:1)
Institute of Applied Physics
University of Tübingen
Auf der Morgenstelle 10
72076 Tübingen
Germany

Joy, David C. (**IV**:2.1)
The University of Tennessee
EM Facility – Programm in Analytical Microscopy
F241 Walters Life Sciences Building
Knoxville, Tennessee 37996/0810
USA

Kruit, Peter (**IX**:1)
Delft University of Technology
Dept. of Applied Physics
Lorentzweg 1
2628 CS Delft
The Netherlands

Mundschau, Michael V. (**VI**:1)
Bowling Greet State University
Dept. of Chemistry
Bowling Green, Ohio 43403-0213
USA

Oleshko*, Vladimir; Gijbels, Renaat (**IV**:2.5)
Department of Chemistry
University of Antwerp (UIA)
2610 Wilrijk-Antwerpen
Belgium

*On leave from the Russian Academy of Sciences
N. N. Semenov Institute of Chemical Physics
117421 Moscow
Russia

Pennycook, Stephen J. (**IV**:2.3)
Solid State Division
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, TN 37831-6030
USA

Scheinfein, Mike R. (**V**:2)
Dept. of Physics and Astronomy
Arizona State University
Tempe, AZ 8528 7-1504
USA

Unguris, John; Kelley, Michael H.; Gavrin, Andrew; Celotta, Robert J.; Pierce, Daniel T. (**V**:2)
Electron Physics Group
Nat. Institute of Standards and Technology
Gaithersburg, MD 20899
USA

Schwarz, Udo D. (**VII**:2)
Institute of Applied Physics
University of Hamburg
Jungiusstr. 11
20355 Hamburg

Spence, John C. H. (**IX**:2)
Dept. of Physics
A. S.U.
Tempe, AZ 85287
USA

Ventrice, Carl A., Jr. (**VII**:4)
Department of Physics
Rensselaer Polytechnic Institute
Troy, NY 12180
USA

Wadas, Andrzej (**VII**:3)
Institute of Applied Physics
University of Hamburg
Jungiusstr. 11
20355 Hamburg
Germany

Wiesendanger, Roland (**VII**:1)
University of Hamburg
Institute of Applied Physics and Microstructure Research Center
Jungiusstr. 11
20355 Hamburg
Germany

Van Espen, Pierre; Janssens, Gert (**IV**:2.6)
Micro and Trace Analysis Centre
Dept. of Chemistry
University of Antwerpen (UIA)
Universiteitsplein 1
2610 Antwerpen
Belgium

Outline

Volume 1: Methods I

Volume 3: Applications

I Classes of Materials

Contents

Volume 2: Methods II

IV Electron Microscopy

V Magnetic Methods

VIII Image Recording, Handling and Processing

IX Special Topics

2 Scanning Beam Methods

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

2.1 Scanning Electron Microscopy

2.1.1 Introduction

The scanning electron microscope (SEM) is the most widely used of all electron beam instruments. It owes its popularity to the versatility of its various modes of imaging, the excellent spatial resolution of its images, the ease with which the micrographs that are generated can be interpreted, the modest demands that are made on specimen preparation, and its 'user-friendliness'. At one end of its operating range the SEM provides images which can readily be compared to those of conventional optical microscopes, while at the other end its capabilities are complementary to instruments such as scanning tunneling (STM) or atomic force (AFM) microscopes. While its resolution can now approach 0.5 nm, rivaling that of a transmission electron microscope, it can handle specimens as large as production size silicon wafers.

The SEM had its origins in the work of von Ardenne [1, 2] who added scanning coils to a transmission electron microscope. A photographic plate beneath the electron transparent sample was mechanically scanned in synchronism with the beam to produce the image. The first recognizably modern SEM was described by Zworykin et al. [3]. This instrument incorporated most of the features of current instruments, such as a cathode-ray-tube display and a secondary electron detector, and achieved a resolution of 5 nm on solid specimens. In 1948 Oatley [4] and his students commenced their work on the development of the SEM leading in 1965 to the first commercial machine the Cambridge Scientific Instruments Mark 1 'Stereoscan'. There are now seven or eight manufacturers of these instruments in Europe, the USA, and Japan, and it is estimated that about 20 000 SEMs are in use worldwide.

The SEM is a mapping, rather than an imaging, device (Fig. 1) and so is a member of the same class of instruments as the facsimile machine, the scanning probe microscope, and the confocal optical microscope. The sample is probed by a beam of electrons scanned across the surface. Radiations from the specimen, stimulated by the incident beam, are detected, amplified, and used to modulate the brightness of a second beam of electrons scanned, synchronously with the first beam, across a cathode ray tube display. If the area scanned on the display tube is $A \times A$ and the corresponding area scanned on the sample is $B \times B$ then the linear magnification $M = A/B$. The magnification is therefore geometric in origin and may be changed by varying the area

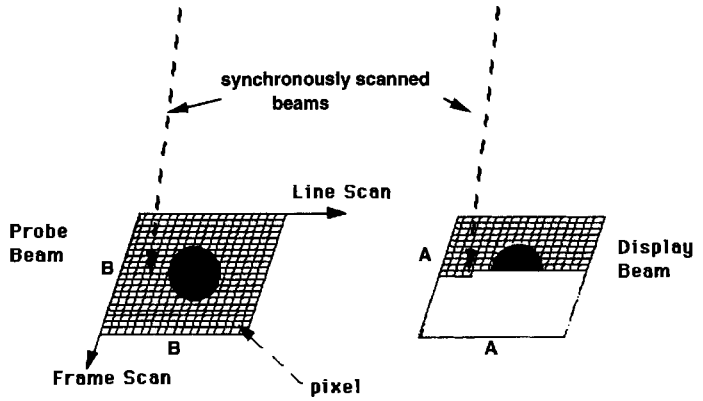


Figure 1. Schematic illustration of the basic mapping principle of the scanning electron microscope.

scanned on the sample. The arrangement makes it possible for a wide range of magnifications to be obtained, and allows rapid changes of magnification since no alterations to the electron-optical system are required. There is no rotation between the object and image planes, and once the instrument has been focused on a given area the focus need not be changed when the magnification is varied. To a first approximation the size of the finest detail visible in the image will be set by the size of the probe scanning the specimen. Multiple detectors can be used to collect several signals simultaneously which can then be displayed individually, or combined, in perfect register with each other. It is this capability in particular which makes the SEM so useful a tool since multiple views of a sample, in different imaging modes, can be collected and compared in a single pass of the beam.

2.1.2 Instrumentation

Figure 2 shows the basic components of the SEM. These can be divided into two main categories, the electron-optical and detector systems, and the scanning, processing, and display systems. The electron-optical components are often described as being the 'column' of the instrument while the other items are the 'console' of the machine. The source of electrons is the gun which produces them either thermionically, from tungsten or lanthanum hexaboride cathodes, or from a field emission source. These electrons are then accelerated to an energy which is typically in the range from 500 eV to 30 keV. The beam of electrons leaving the gun is then focused on to the specimen by one or more condenser lenses. Although either electrostatic or electromagnetic lenses could be employed all modern SEMs use electromagnetic lenses. Typically, the final objective lens has been of the pin-hole design with the sample sitting outside the magnetic field of the lens since this arrangement gives good physical access to the specimen. However, in this arrangement the specimen is 10 to 20 mm away from the lens which must therefore be of long focal length and correspondingly high aberration coefficients. In modern, high performance, instruments it is now common to use an immersion lens [5], in which the sample sits inside the lens at the center of

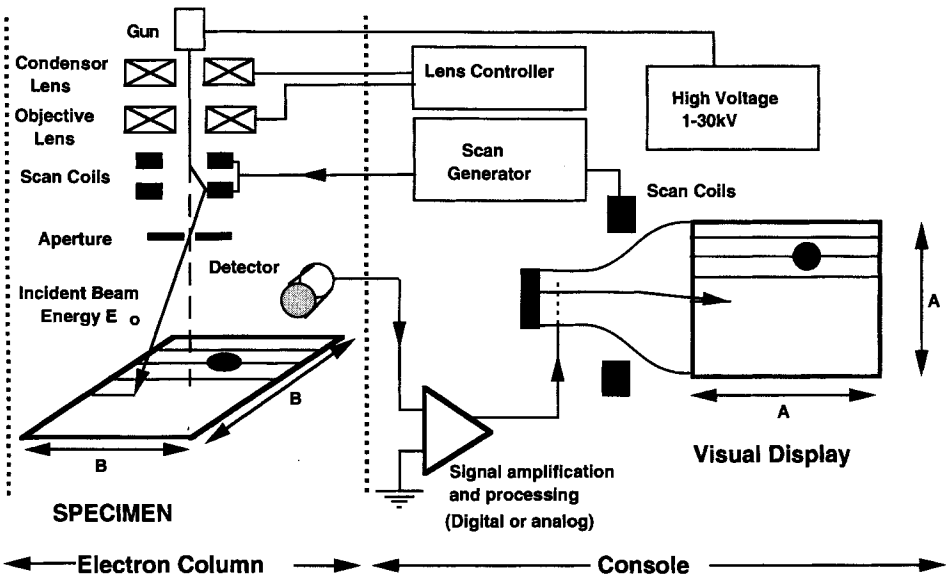


Figure 2. Basic components of the scanning electron microscope.

the lens field, or a 'snorkel' lens [6] in which the magnetic field extends outside of the lens to envelope the sample. Although the immersion lens gives very good performance and, by making the sample part of the lens structure, ensures mechanical stability, the amount of access to the specimen is limited. The snorkel lens on the other hand combines both good electron-optical characteristics with excellent access for detectors and stage mechanisms.

The coils that scan the beam are usually incorporated within the objective lens. A double scan arrangement is often employed in which one set of coils scans the beam through some angle θ from the axis of the microscope while a second set scans the beam through an angle 2θ in the opposite direction. In this way all scanned beams pass through a single point on the optic axis allowing for the placement of a defining aperture without any constriction of the scanned area. The scan pattern, or 'raster', produced on the specimen, is usually square in shape and is made up of 1000 horizontal lines each containing 1000 individual scanned points or 'pixels'. The final image frame thus contains 10^6 pixels, although for special activities such as focusing or alignment frames containing only 256×256 pixels may be used. A variety of detectors are provided in the SEM, including an Everhart–Thornley [7] scintillator detector for secondary electrons, some type of detector for backscattered electrons and, often, a detector for fluorescent X-rays to permit chemical microanalysis with the beam. Further details of these devices are given below.

Signals from any of the detectors are amplified and presented to the display screens in the console. The electronics provide control of the amplification, DC offset, and bandwidth of the signal. Increasingly the detector output is passed

through an analog to digital converter (ADC) and then handled digitally rather than as an analog video signal. This permits images to be stored, enhanced, combined, and analyzed using either an internal or an external computer. Although the majority of images are still recorded on to photographic film, digital images can be stored directly to magnetic or magneto-optic discs and hardcopy output of the images can then be obtained using laser or dye-sublimation printers. The scan drive currents to the scan coils may also be digitized to provide precision control of the beam position although analog scans are still most commonly encountered. Typically scan repetition rates ranging from 15 or 20 frames per second (TV rate) to one frame in 30 to 60 s (photographic rate) are provided. In addition individual pixels, or arrays of pixels, within an image field may be accessed if required.

2.1.3 Performance

It is usual to define the performance of a microscope in terms of the spatial resolution that it can produce. In the case of the SEM the attainable resolution is determined by a number of factors including the diameter d of the electron-beam probe that can be generated, the current I_b contained in that probe, the magnification of the image, and the type of imaging mode that is being used. Over most of the operating energy range (5–30 keV) of the SEM the probe size and beam current are related by an expression of the form [8]

$$d = (C_s^{1/4}\lambda^{3/4})\left[1 + \frac{I_b}{\beta\lambda^2}\right]^{3/8} \quad (1)$$

where λ is the wavelength of the electrons ($\lambda \approx 1.226E_0^{-1/2}$ nm where E_0 is the incident electron energy in eV), β is the brightness of the electron gun in A cm^{-2} steradian^{-1}, and C_s is the spherical aberration coefficient of the objective lens. Equation (1) contains two groups of terms; the first of which ($C_s^{1/4}\lambda^{3/4}$) can be taken as being the minimum spot size that the lens can produce; the second term then multiplies this limiting probe size by a factor which depends on the current I_B required in the beam. Note from Eq. (1) that the obtainable spot size d is always larger than the minimum value except in the limiting case when I_B is zero, so for any finite beam current the probe size must be enlarged and, hence, the spatial resolution of the image will be worsened. The degree to which this occurs will depend both on the current I_B required and on the gun brightness β. Figure 3 plots the value of this multiplying factor as a function of gun brightness for two beam currents, firstly $I_B = 10 \times 10^{-12}$ A, a typical value for high resolution secondary electron imaging, and secondly $I_B = 1 \times 10^{-9}$ A, a value suitable for backscattered imaging, electron channeling, or other special modes of imaging. For a brightness of 10^5 A cm^{-2} sr^{-1}, such as encountered with a tungsten cathode thermionic gun [9], the multiplying factor is 5.6 for imaging currents and over 30 for the higher current case. In such a case the imaging performance of the SEM is dominated not by the lens but by the relatively poor performance of the electron gun. If the brightness is increased to 10^6 A cm^{-2} sr^{-1}, for example, by using a lanthanum hexaboride cathode in the gun [9], then the multiplying factors become 2.5 and 1.3, respectively, showing that improving the performance of the gun

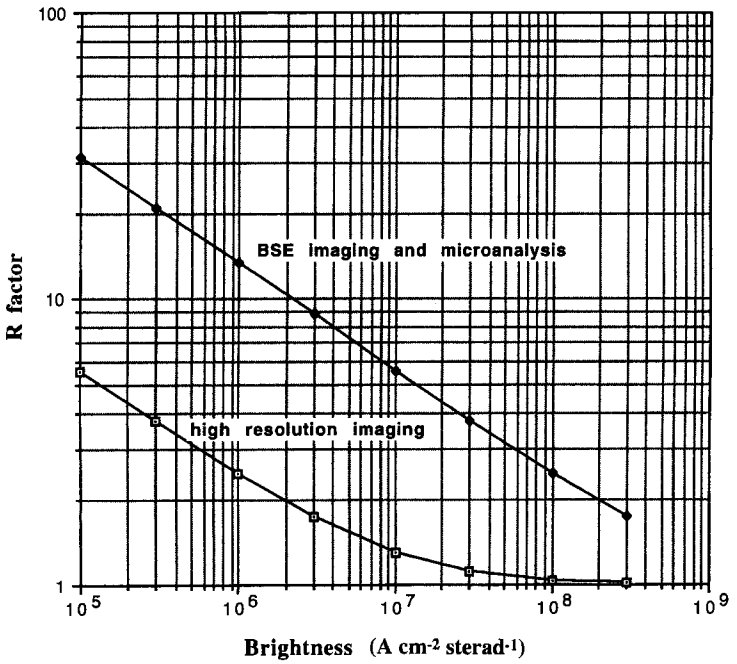


Figure 3. *R* factor modifying minimum probe size against brightness for high resolution SE imaging ($I_B = 10^{-11}$ A) and backscattered imaging or microanalysis ($I_B = 10^{-9}$ A).

dramatically enhances the resolution of the microscope. Finally, if the gun brightness is further increased to 10^8 A cm^{-2} sr^{-1} by using a field emission source [9] then the factor is close to unity for both modes of operation considered. In this case the probe forming performance is no longer limited by the brightness of the source but is controlled by the properties of the lens. (Note that since electron optical brightness β increases linearly with beam energy E_0, while the electron wavelength λ varies as $E_0^{-1/2}$, the value of the quantity $\beta\lambda^2$ is independent of the energy actually used.) For a modern SEM C_s is typically a few millimeters; thus minimum probe sizes of 1 or 2 nm are available. At low beam energies (below 5 keV) additional effects including the energy spread of electrons in the beam must also be considered, but the general conclusions discussed above remain correct.

A second limitation to the performance comes from the fact that the scan raster divides the image into a finite number of pixels, typically 1000 lines each of 1000 pixels. If the size of the display CRT is $A \times A$, and the instrumental magnification is M then the size of each pixel referred to the sample is $A/(1000M)$. For example, if A is 10 cm and $M = 20$, the effective pixel size on the sample is then 5 μm. Each pixel represents a single piece of information in the image, so no detail smaller than the pixel size can be resolved even though the probe size may be significantly smaller than this value. At low magnifications the SEM is therefore pixel limited in its resolution; for example, it is not until a magnification of 10 000× that the pixel size falls below 10 nm and so becomes comparable in size to the probe diameter discussed above.

Finally, the spatial resolution of an image may be limited by the lateral spread

of the electron interactions that produce the desired image information. The extent of this varies widely with the mode of operation and with the nature of the specimen, from a nanometer or less for secondary electron imaging in some circumstances, to a micrometer or more for backscattered or EBIC imaging. In summary, therefore, the resolution of the SEM cannot be defined by a single number but depends on many different factors. In the past the performance of the SEM has been mainly limited by gun brightness, lens quality, and other instrumental parameters, but with current instruments employing field emission guns and immersion lenses fundamental electron interaction effects now probably define the performance more than any specific attribute of the microscope itself.

2.1.4 Modes of Operation

2.1.4.1 Secondary Electron Imaging

Secondary electrons (SE) are those electrons emitted by the specimen, under irradiation by the beam, which have energies between 0 and 50 eV. Because of their low energy the SE only travel relatively short distances in the specimen (3–10 nm) and thus they emerge from a shallow 'escape' region beneath the surface. There are two cases in which an SE can be generated and subsequently escape from the specimen: first, when an incident electron passes downwards through the escape depth, and secondly, as a backscattered electron leaves the specimen and again passes through the escape region. Secondary electrons produced in the first type of event are designated SE1 and, because they are generated at the point where the incident beam enters the specimen, it is these which carry high resolution information. The other secondaries are called SE2, and these come from a region whose size is of the order of the incident beam range in the sample. Since this can be 1 μm or more at high energies and it can be seen that the SE2 carry low resolution information. The SE1 and SE2 signals cannot be separated by any device because they are identical in their properties but when the SEM is operated at high magnifications the area scanned by the beam is less than the area from which the SE2 signal is generated so the SE2 signal is effectively independent of the beam position and forms a constant background to the SE1 signal. The variations in the SE1 signal can then be isolated from the SE2 variations, although since the SE1 component is only about one-half to one-third as strong as the SE2 component this means that the high resolution information is diluted by the low resolution background. The yield δ of secondary electrons (δ = number of SE per incident electron) varies with the energy of the incident electron beam. At high energies (10 keV or more) the yield is typically only 0.1 to 0.2, but as the beam energy is reduced the yield rises rapidly and may exceed unity for energies of the order of 1 keV. SE imaging is, therefore, preferentially performed at a low rather than a high accelerating voltage.

Secondary electron imaging is the most common mode of operation of the SEM and it has been estimated that 95% of all published SEM images have been recorded with the SE signal. The importance and utility of this mode is the result

of several factors:

(i) Secondary electrons are easy to collect.
(ii) Secondary electrons carry information about the surface topography of the specimen. Information about surface chemistry, and magnetic and electric fields may also be obtainable on suitable specimens.
(iii) SE images can, in most cases, be interpreted readily without specialist knowledge.
(iv) The SE image can display information at a spatial resolution of 1 nm or better under properly optimized conditions.

The practical key to the success of SE imaging has been the detector originally described by Everhart and Thornley [7], shown in Fig. 4. The secondary electrons are allowed to strike a scintillator material after having been accelerated to an energy of about 10 keV by a positive potential applied to the front face of the detector. To prevent this potential deflecting the incident electron beam the scintillator is often surrounded by a Faraday cage biased to about +200 V to attract the secondaries and made of open metal mesh. The light produced in the scintillator, which is usually either a plastic or a rare-earth doped YAG crystal, is transferred down a light-pipe made of quartz and leaves the vacuum chamber of the SEM through a window. On the other side of the window the light enters a photomultiplier tube where it is reconverted to an electronic signal. This arrangement produces a very high amplification of the original SE signal while only adding insignificant amounts of noise, a wide dynamic range (because of the logarithmic characteristic of the photomultiplier tube), and a rapid response to changes in the signal intensity. It is cheap

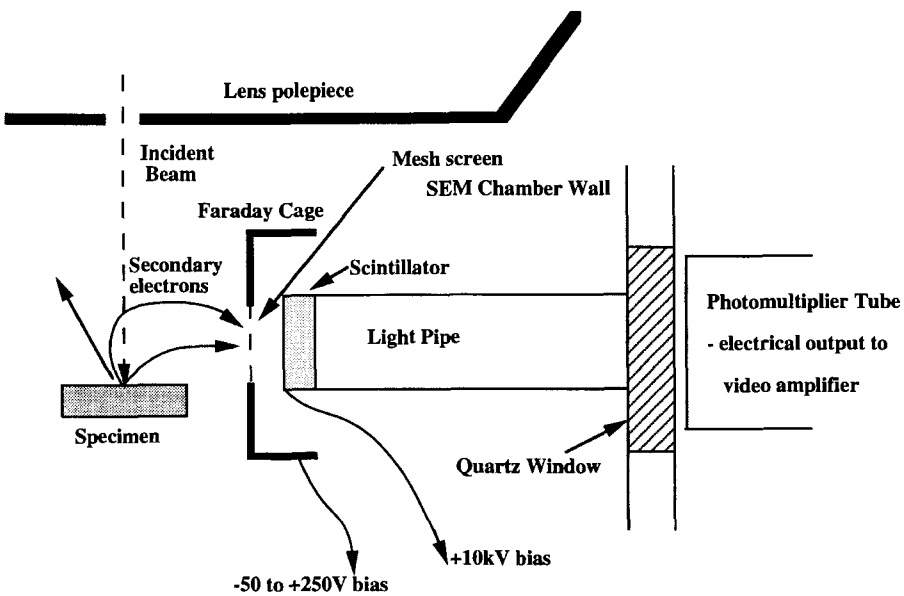


Figure 4. Everhart–Thornley SE detector and specimen chamber geometry.

to produce, and the detector may be placed anywhere within the specimen chamber of the SEM and still collect 50 to 60% of the SE leaving the specimen. On some advanced instruments the SE detector may be placed above the objective lens and out of line of sight of the sample [10]. The magnetic field of the lens then collects the SE which pass back through the lens and are then extracted by the detector. This arrangement provides good collection efficiency and is well suited for high resolution imaging.

The general characteristics of SE signals are illustrated in Fig. 5 which shows a collection of diatoms. The image has a marked three-dimensional effect caused by the lighting and shadow effects which decorate each object, and each edge is outlined by a fine bright line which adds definition to the detail. The ability to view topographic detail in the image occurs because the yield of secondary electrons varies with the angle between the incident electron beam and the local surface normal [9]. Areas which are at a high angle to the beam are bright (large signal) compared to those faces which are normal to the beam (small signal), and faces looking towards the detector (which is in the top right hand corner of the micrograph) are in general brighter than those facing away from the detector although, because the SE are readily deflected by the electrostatic field from the detector, both faces are easily visible. This type of behavior is analogous to that of visible light as described by Lambert's cosine law [9]. If a source of light were placed on the SE detector then an observer looking down from the electron gun would see the brightest illumination on those surfaces tilted towards the light, and less light on those surfaces facing upwards towards the observer. Using this analogy it is easy to interpret secondary electron images in a reliable and consistent way. The bright edges in the image occur because secondary electrons can escape through two faces, rather than just one, in the vicinity of an edge and therefore the signal is anomalously high in such regions.

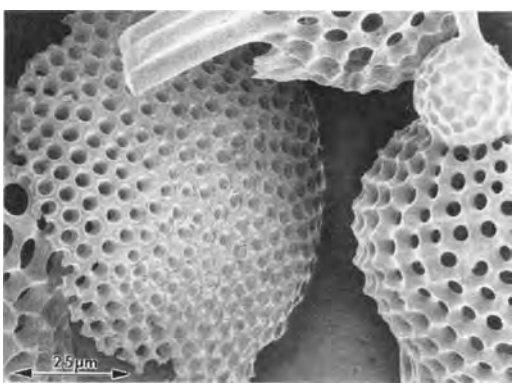


Figure 5. Secondary electron images of Radiolarium. Recorded in Hitachi S-4500 field emission SEM at 5 keV beam energy. Magnification: 800×.

Even at much higher magnifications the appearance and interpretation of the SE image remain much the same. Figure 6 shows an image of the magnetic media on the surface of a computer hard disc. Although the detail is now only of a few nanometers in scale the image and its relationship to the surface topography can be understood in the same way as that described above. The fact that SE images can be consistently and easily interpreted over a wide range of magnifications has been a significant factor in the popularity of the SEM. This is in marked contrast to the situation on, for example, a scanning tunneling microscope (STM) which has similar spatial resolution but for which the multiplicity of possible effects contributing the image is such that

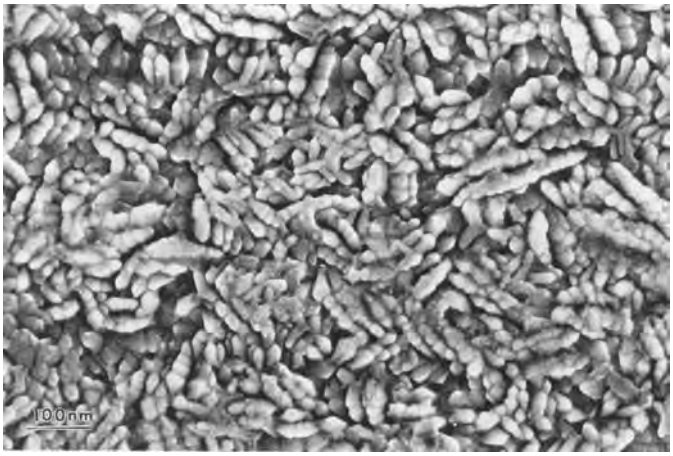


Figure 6. High resolution image of magnetic disc media surface recorded at 30 keV in Jeol JSM890 field-emission SEM.

interpretation is difficult and unreliable. At the highest resolutions (below 1 nm) the details of the electron–solid interactions that produce the SE must be considered and image interpretation becomes more complex and research is still in progress in this area [12].

Secondary electron images can also carry information about other properties of the specimen. One such mode is voltage contrast which is illustrated in Fig. 7. The micrograph shows a heart pacemaker chip, powered up and running but removed from its usual protective casing, which is being examined in the SEM in the SE mode. In addition to the normal topographic contrast, large scale regions of uniform bright and dark contrast are visible. The bright areas are those which have a potential which is negative with respect to ground, while the dark areas are positive with respect to ground. The origin of such contrast is straightforward. When an area is negative then the collection field from the SE detector is increased and a higher fraction of the SE are collected. An area that is positive experiences a lower collection field from the detector, and also

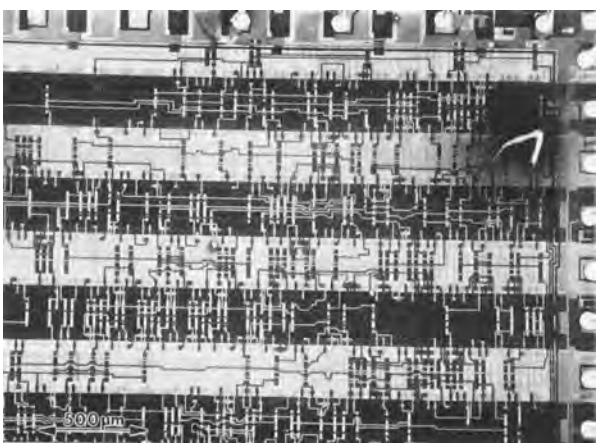


Figure 7. Voltage contrast from integrated circuit. Recorded at 5 keV in Hitachi S-800 FE SEM.

has a tendency to recollect its own secondary electrons so the SE signal from such areas is lower. In this simple form the technique gives a qualitative view of static (DC) potential distributions but, by improvements in instrumentation, it is possible to study potentials which may be varying at frequencies up to 100 MHz or more [13], and to measure the potentials with a voltage resolution of ±10 mV and a spatial resolution of 0.1 μm. This unique ability of the SEM to measure voltages in real-time, from small areas, and without requiring any mechanical contact has been of considerable value in the development of semiconductor technology and represents a significant fraction of the usage of this instrumentation.

Contrast can also be generated by the presence of magnetic fields outside of a sample, for example, from the leakage fields that exist above the surface of a uniaxial magnetic material such as cobalt. In this case the contrast is produced by the Lorentz deflection of the SE after they leave the specimen. A field in one direction will deflect more SE towards the detector, while a field in the opposite sense will deflect the SE away from the detector [9]. In a more sophisticated application of this approach the incident electron beam is polarized, and the change in polarization of the emitted SE, measured by special detectors, is used to produce the signal contrast [14].

2.1.4.2 Backscattered Electrons

Backscattered electrons (BSE) are defined as being those electrons emitted from the specimen which have energies between 50 eV and the incident beam energy E_0. Unlike the secondary electrons which are produced as the result of the incident electron irradiation, backscattered electrons are incident electrons which have been scattered through angles approaching 180° within the sample and consequently leave the sample again. The yield η of BSE (η = number of BSE per incident electron) varies monotonically with the atomic number Z of the specimen, increasing from about 0.05 for carbon to about 0.5 for gold. At high incident beam energies, therefore, the number of BSE produced is greater than the number of SE but despite this fact backscattered electron imaging has received less attention and use than SE imaging until recently. This is because of the practical problem of efficiently collecting the BSE. Since the energy of the BSE is of the order of $E_0/2$ they are much more difficult to deflect towards a detector than the SE. Consequently the detector must be placed in a suitable position above the specimen to intercept the BSE, and must be physically large enough to collect a high fraction of the signal. Several successful detector designs are now in current use, including scintillator systems similar to the Everhart–Thornley SE detector discussed above, solid-state detectors, and electron-multiplier devices such as the microchannel plate.

Because the yield of BSE varies with the atomic number of the specimen the most widespread use of backscattered electrons is atomic number, or Z-contrast, imaging. Figure 8 shows a backscattered image of a sample of a 5000 year old Assyrian glass. Although the SE image showed the material as being homogeneous, the backscattered image reveals the presence of numerous precipitates and stringers

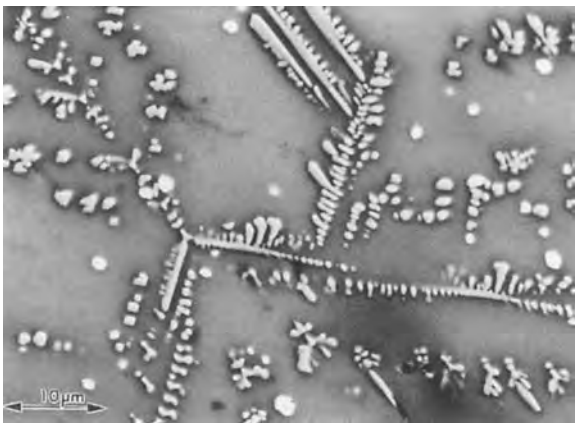


Figure 8. BSE *Z*-contrast image of Assyrian glass showing Cu_2O dendrites. Recorded on Cambridge S250 SEM at 10 keV.

which were later determined to be Cu_2O, probably dissolved from the vessel in which the glass was melted. The phases are readily distinguished here because the glass, essentially SiO_2, for which the mean atomic number is 10, produces significantly fewer backscattered electrons than do the precipitates for which the mean atomic number is 22. In cases such as this BSE imaging provides a quick and convenient method of examining the distribution of chemistry within a material and qualitatively separating regions of high and low atomic number. The technique can also be used to examine diffusion gradients across a boundary, to scan for the presence of unexpected contaminants, and to look for evidence of multiphase structures. Atomic number contrast imaging has also been widely applied in biology [15] by attaching high atomic number elements such as gold to active molecular groups which preferentially bind at specified locations on a cell. A backscattered image of the material then shows the heavy metals as bright spots against the dark background of the predominantly carbon matrix. By comparing the BSE image with the SE image the binding locations can be unambiguously located.

Despite several commercial attempts to justify the procedure, *Z* contrast imaging cannot be used as a substitute for proper microanalytical techniques. First, for all but pure elements, there are an infinite number of ways in which the same average atomic number can be achieved by combining different materials. Second, although it is widely assumed that the variation of backscattering yield with mean atomic number is monotonic for a compound there is no solid experimental evidence proving this and there are, in fact, examples where anomalous behavior has been reported [16]. Third, BSE detectors respond not only to the number of backscattered electrons but also to their energy and as a result two materials giving the same backscattered signal may have different backscattered yields compensated by the difference in the mean energy of the BSE. The technique should thus be regarded as a valuable diagnostic and observational mode but should not be used quantitatively without very careful preparation and calibration [17].

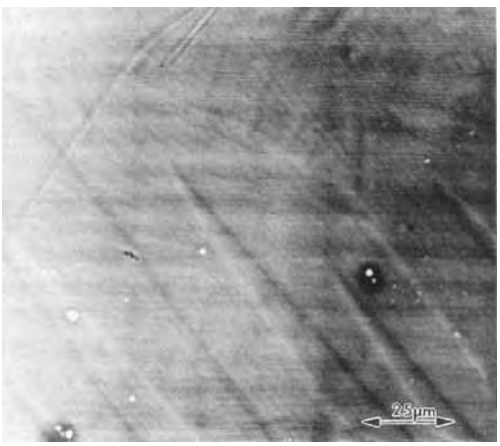


Figure 9. Magnetic contrast from domain walls in Fe-3%Si transformer core material. Recorded on Cambridge S250 at 30 keV.

The backscattering coefficient of a material can also be modified by other factors, including its magnetic configuration [18]. Figure 9 shows the backscattered image from Fe–3%Si transformer core material which has cubic magnetic anisotropy. The characteristic 'fir tree' magnetic domain structure of this type of material is clearly visible outlined by the black and white lines crossing the micrograph. The contrast in this case arises because of the Lorentz deflection of the incident beam within the specimen in the vicinity of the domain boundaries. For some directions of the magnetic flux across the domain the incident beam suffers a small additional downwards deflection which lowers the backscattering coefficient, while for flux in the opposite sense the incident electrons are deflected slightly upwards towards the surface and the backscattering yield is increased.

Unlike secondary electrons which come from the surface region of the specimen, backscattered electrons emerge from a volume which is of the order of one-third of the incident beam range in depth and radius. To a first approximation, the range R (in nm) is given by

$$R = \frac{75E_0^{5/3}}{\rho} \tag{2}$$

where E_0 is the beam energy in keV and ρ is the density of the specimen in g cm^{-3}. Thus for beam energies of 10 keV, and typical densities, the backscattered signal comes from a region of the order of a fraction of a micrometer or more in breadth and depth. Consequently, the backscattered image contains information about the interior of the specimen rather than the surface but at the expense of the spatial resolution which is generally lower than that of the corresponding SE signal. For example, backscattered images can reveal the presence of buried voids or cracks in materials that are not visible at all in the SE image.

Unique to the backscattered image is information related to the crystalline nature of the sample. The origin of this contrast is illustrated schematically in Fig. 10. If the incident electron beam enters a crystal at a random angle of incidence then backscattering will occur in the normal way, and with the usual yield. But if the incident beam is aligned along a symmetry direction of the lattice then the incident electrons tend to channel between the lattice planes and as a result they penetrate more deeply into the specimen and the backscattering yield is reduced. Although this particle model is not physically realistic a detailed analysis using dynamical diffraction theory [19] confirms the existence of such a phenomenon. Thus if the angle of incidence between the electron beam and a crystal is varied then the backscattering yield is modulated at angles corresponding to the symmetry directions

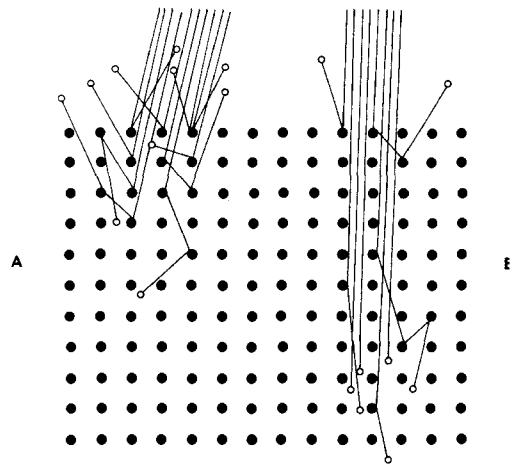


Figure 10. Schematic illustration of the origin of electron channeling contrast (a) random incidence direction, (b) incidence along a symmetry direction.

of the crystal. The image, or electron channeling pattern, produced in such a case is illustrated in Fig. 11 for a single crystal of InP. Here the variation in the angle of incidence has been achieved simply by scanning the large crystal at low magnification, so that between extremes of the field of view the angle between the beam and the surface normal varies by about ±8°. The

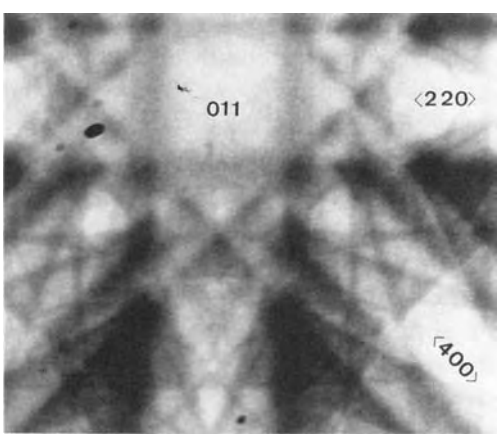


Figure 11. Electron channeling pattern from InP recorded at 25 keV on Cambridge S250 SEM.

micrograph therefore contains both spatial and angular information. The small circular features scattered over the micrograph are conventional images of dust particles on the surface, while the larger scale linear structures are crystallographic in origin and are related to the angle of incidence. The broad bands crossing the micrograph horizontally, vertically, and diagonally, are the channeling contrast from lattice planes in the {2 0 0} and {2 2 0} type zones in the crystal. These bands cross with four-fold symmetry about a point whose angular position marks the direction of the (0 1 1) pole of the crystal. Moving the crystal laterally will not change its symmetry (unless some crystallographic boundary is crossed) and hence the channeling pattern will not change, but if the crystal is tilted or rotated then the pattern moves as if rigidly fixed to the lattice. Note that in the channeling condition the observed contrast comes only from regions within two or three extinction distances of the surface (i.e., typically 30 to 50 nm at 20 keV) not from the full depth from which the BSE can emerge [19]. The quality of the crystal in the surface region is therefore important, and samples must be chemically or electrochemically polished to give good patterns.

The geometry and appearance of the channeling pattern is similar to that of a Kikuchi pattern in transmission electron microscopy (i.e., a gnomic projection) and contains similar information about the crystal. The angular width of the bands is twice the appropriate Bragg angle for the given lattice spacing and electron wavelength and so for a known accelerating voltage lattice spacings may be deduced from the measured width after the angular scale of the pattern is obtained by using

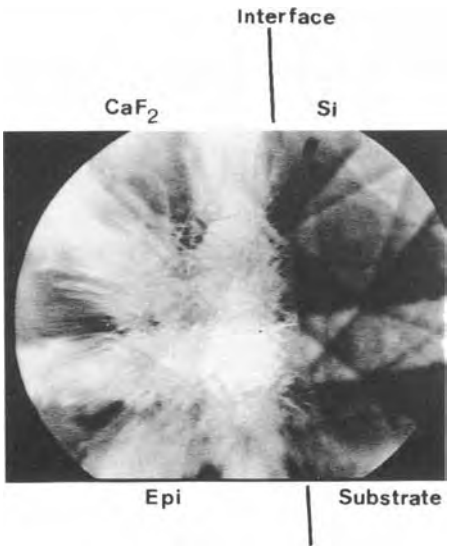


Figure 12. Selected area channeling pattern across the epitaxial interface between CaF_2 and silicon. Recorded on Cambridge S250 at 25 keV.

a crystal of known spacings. From the lattice spacings and interplanar angles the pattern can then be indexed, and the symmetry, orientation and space group of the crystal can be determined.

By a modification of the scan arrangement it is possible to obtain selected area channeling patterns containing information from areas as small as 1 μm in diameter [19]. This is achieved by allowing the beam to rock about a fixed point on the surface but not scan laterally. Figure 12 shows an example of the utility of this arrangement. The sample is of calcium fluorite CaF_2 grown epitaxially on silicon and the sample has been cross-sectioned to reveal the interface. The continuity of the selected area channeling pattern across the interface region confirms that epitaxy has been achieved but a comparison of the patterns on the two sides of the interface shows that while the pattern from the silicon contains sharp, well defined, detail the corresponding pattern from the CaF_2 is much less well defined. Although the major features of the pattern are still discernible the fine detail is absent. This effect is attributable to the presence of dislocations, and hence of strain, in the CaF_2 layer. The variations in lattice orientation around each dislocation have the same effect as superimposing two patterns of slightly different orientation. Consequently the detail in the pattern becomes broader and less distinct and, at high enough dislocation densities, disappears completely. The channeling pattern can therefore be analyzed to determine the quality of the crystal as well as its other attributes. This technique has been successfully applied to studies of deformation, annealing, superplasticity, and to the study of stress concentrations associated with corrosion cracking [19].

In the limit the technique can be used to visualize individual defects within the crystal as shown in Fig. 13. If the crystal, here a bulk sample of MoS_2, is oriented with respect to the beam so as to avoid any strong reflections then the lattice distortions which occur around a dislocation may be sufficient to rotate the lattice into a strong channeling condition and so produce visible contrast as shown in the micrograph. Each of the bright lines represent a single dislocation threading through the crystal. In this particular example several separate sets of defects are evident. Such images obey the same rules as diffraction contrast micrographs in the TEM and so the Burger's vector of dislocations may be determined by imaging in several different orientations [20]. Because specimens need not be thinned to produce dislocation contrast in this mode, and because the specimen as a result has only

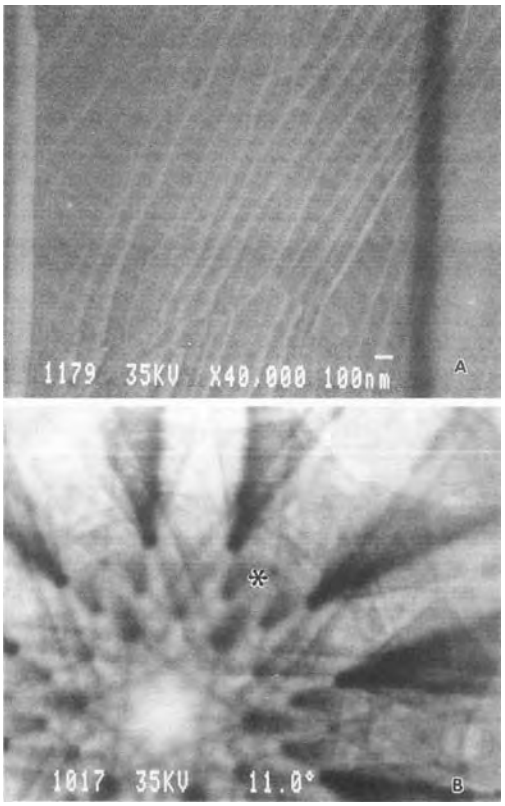


Figure 13. (a) Defect image in MoS_2 recorded at 25 keV on Jeol JSM890 FM SEM. The star on the channeling pattern (b) indicates the orientation to which the crystal was set for the micrograph.

one rather than two free surfaces, this technique is more rapid than the corresponding TEM methods and uses a specimen which is more physically realistic. It can also be usefully employed even at very low dislocation densities where the chance of finding a defect in the field of view of a TEM would be very small.

2.1.4.3 Special Techniques

Unlike the techniques already discussed the imaging modes that will now be discussed are specific to a particular class of materials, those which are semiconductors. Because of the technological and economic importance of these materials the capabilities of the SEM in this field have proven to be of great value.

A semiconductor is a material, Fig. 14, in which the valence band and the conduction band energy levels are separated by a band gap. Depending on the position of the Fermi energy level the conduction band is therefore either empty of electrons, or contains only a very small number and the material has little or no electrical conductivity. If a fast electron is injected into the semiconductor then some of the energy that it deposits in the specimen can be used to promote electrons across the band from the valence band to the conduction band. Since the valence band was initially completely filled the removal of an electron also produces a hole in this band. Because the electron is negatively charged and the hole has a positive charge they drift through the material together and so it is convenient to describe them as being an electron–hole pair. The energy E_{eh} required to form one electron–hole pair is about three times the band gap, for example, in silicon E_{eh} is 3.6 eV. A single incident electron of energy E_0 can then create about E_0/E_{eh} electron–hole pairs, that is approximately 3000 pairs for each 10 keV incident electron.

In the absence of any external stimulus the electrons and holes will drift through the lattice, staying physically close to each other so as to maintain overall electrical neutrality, and within a short time, typically 10^{-15} to 10^{-12} s after the initial excitation each electron will drop back into a hole releasing its excess energy in one or more ways including the production

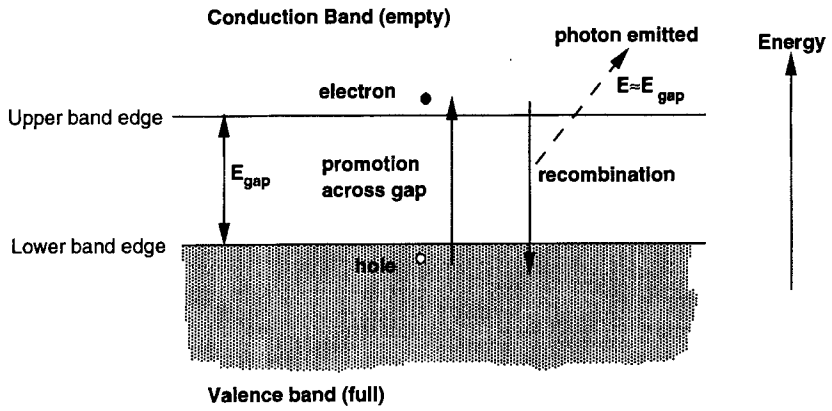


Figure 14. Band gap diagram for a semiconductor.

of cathodoluminescence (CL) which is discussed below. However, a semiconductor – unlike a conductor – can sustain a potential difference across itself. If a voltage is applied across the material then the resultant electric field will cause the holes to move towards the negative potential and the electrons towards the positive potential. This motion of charge carriers constitutes a current flow and hence the incident electron beam has produced conductivity in the semiconductor. If the incident beam is turned off the current flow will cease. This phenomenon is therefore referred to as electron beam induced conductivity (EBIC).

(a) *Charge Collection Microscopies (EBIC)*

A field can be produced across a semiconductor in three ways as shown schematically in Fig. 15. The first way, Fig. 15a, is to apply a potential from an external source. This mode is often referred to as β-conductivity and has not found widespread use because it offers no advantages over either of the other techniques available [13]. The first practical technique for employing electron–hole pair carriers as a signal source is that shown in Fig. 15b which uses the depletion field which exists around a p–n junction in a semiconductor. In the arrangement shown the junction is short circuited through the external conductor. With no incident beam of electrons no current flows, but there is a potential difference between the p and n sides of the junction of typically 0.5 to 1 V. The field associated with this potential difference extends for a distance which depends on the resistivity ρ of the material but which is typically a few micrometers on either side of the physical position of the junction. This region is called the depletion zone because it can contain no mobile charge carriers. If the electron beam is allowed to fall on to the semiconductor far away from the junction then, although electron–hole pairs are being generated, they are in a field-free region and so will recombine without any net charge flow. But if the beam is placed within the depletion region then the field will separate the electrons and holes, a net motion of charge will occur, and a current I_{cc} will flow around

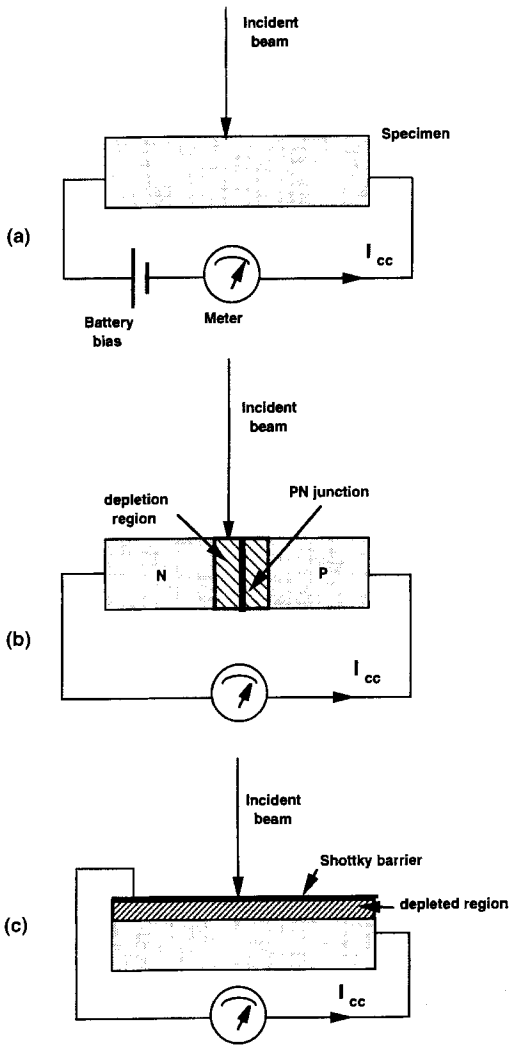


Figure 15. Schematic illustration of three ways of performing EBIC: (a) external battery bias; (b) p–n junction; or (c) Schottky surface barrier.

the external circuit. This current will be $e(E_0/E_{eh})$ where e is the electronic charge and is referred to as the electron beam induced current (EBIC). (To avoid confusion it is better to describe all of these modes as being 'charge collected microscopies'.) If the beam is placed outside of the junction region but close to it then some of the electrons and holes will diffuse into the depletion region, be separated, and generate a signal. If the distance between the incident beam and the junction is X then the fraction $f(X)$ of the carriers which diffuse to the junction is

$$f(X) = \exp(-X/L) \tag{3}$$

where L is the minority carrier diffusion length (i.e., the diffusion length for the holes in the n-type material, or for the electrons in the p-type material). The charge collected signal I_{cc} therefore falls away exponentially on either side of the junction at a rate dependent on the value of L, which is typically 1 to 10 μm for common semiconductors.

This behavior was first observed experimentally by Everhart and Wells [21] and provides a convenient and powerful way of examining the electrically active regions of semiconductor devices. An integrated circuit can be used to produce an EBIC image by connecting a suitable amplifier, that is, one that is capable of responding to currents in the 1 nA to 1 μA range, across the + and − power lines into the device. Figure 16 shows an image formed in this way from a shift register device, together with the corresponding SEI image. Junctions from all of the transistors and diodes in the field of view and within the penetration depth of the incident beam contribute to the observed contrast in the EBIC image; this technique is therefore widely applied when reverse engineering a chip as it provides a detailed layout of the structure of the device. Two extremes of the signal are visible, bright white and dark black, corresponding to current flowing from either a p–n or an n–p junction since in these two cases I_{cc} has the same magnitude but opposite sign. None of the topographic contrast visible in the SE

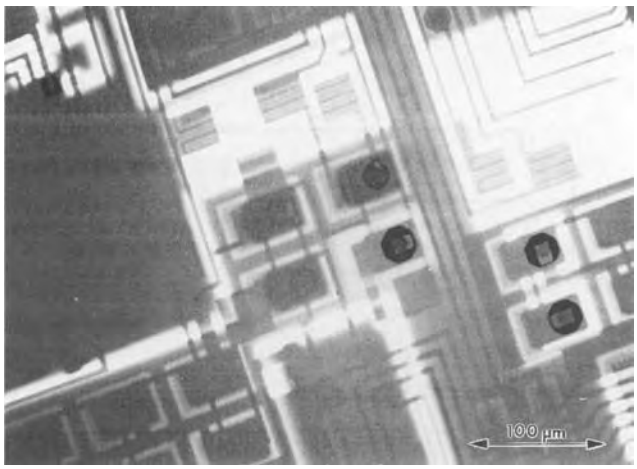


Figure 16. EBIC image of p–n junctions in an integrated circuit. Recorded at 15 keV in Hitachi S-800 FE SEM.

image appears in the EBIC image because the collection of the electron–hole pair signal is dominated by the internal electric fields of the device rather than by anything happening at the surface. At the magnification with which this image was recorded the edges of the p and n regions look sharp, but if the magnification were to be increased then the boundaries would be found to be blurred because of the diffusion of the electrons and holes in the semiconductor. Although this sets a limit to the spatial resolution of the EBIC image it is also a valuable tool because if the variation of I_{cc} with beam position is measured then by using Eq. (3) the minority carrier diffusion length L can be deduced. Since L is greatly affected by the processing that the semiconductor has undergone the ability to measure the diffusion length on a micrometer scale provides an invaluable diagnostic tool in device fabrication, failure analysis, and quality assurance testing.

The limitation of using a p–n junction to collect the electron–hole pairs is that efficient collection of the signal only occurs within a few micrometers on either side of the junction and its depletion region. The technique is made more useful, as shown in Fig. 15(c), when a Schottky barrier is deposited on to the surface of the semiconducting material. The depleted region now extends beneath the whole area of the barrier, and downwards from the surface to a depth dependent on the resistivity of the material. By depositing a barrier several millimeters in size on to a semiconductor large areas of the material can therefore be examined. The Schottky barrier is a metal film, usually titanium or chromium, evaporated on to the atomically clean surface of the material [13] and acts in the same way as a p–n junction. The EBIC signal is collected by making electrical contacts to the barrier itself and to the semiconductor. The benefit of this approach is that it can be used to look at a material before it is processed into a device. If the material were perfect then the EBIC image would be of uniform brightness. But any factors which modify the electrical characteristics of the semiconductor will produce visible contrast effects. Figure 17 shows the EBIC image from a GaAs wafer. The circular area is

Figure 17. EBIC image of doping variations in GaAs wafer. Recorded in Cambridge S250 SEM at 15 keV.

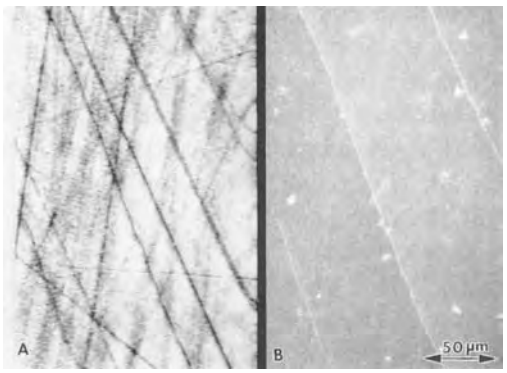


Figure 18. (a) EBIC image of individual crystallographic defects in thin film of Si regrown by hot-wire over an SiO_2 layer. (b) Corresponding SE image showing the linear grain boundaries. Images recorded in Cambridge S250 at 15 keV.

the outline of the Schottky barrier region, and the shadow of the electrical contact on to the barrier is also visible. The variations in brightness across the material are due to impurities in the wafer. Although these are only present at a concentration of $10^{16}\,cm^{-3}$ (i.e., 1 part in 10^7) they vary the resistivity of the material and hence the depletion depth beneath the barrier. If the range of the incident electron beam [Eq. (2)] is greater than the maximum depletion depth then an increase in the depletion depth will increase the signal collected, and vice versa. The extreme sensitivity and speed of this technique makes it ideal for the characterization of as-grown semiconductor crystals [22].

Electrically active defects in a crystal also produce contrast in the EBIC image, as shown in Fig. 18. In this micrograph from a wafer of silicon deformed 0.4% at 670 °C a network of dark lines can be seen. Each line is the trace of an individual dislocation in the material. Because of the presence of dangling bonds at the core of the defect, the dislocation is electrically active and acts as a recombination site. If the electrons and holes recombine at the defect they do not contribute to the external measured current I_{cc} and so the signal intensity falls. The width of the defect line image depends on the local diffusion length in the material as well as on the depth beneath the surface at which the defect is lying, but it is typically a fraction of a micrometer. The ability of the SEM to image defects in a bulk wafer in this way is of great importance, especially in modern semiconductor materials where the defect density may be only 1 to $100\,cm^{-2}$. Conventional techniques such as transmission electron microscopy can only examine areas of the specimen a few micrometers in diameter, and the chance of finding a defect within such a small area is very low. Because the EBIC defect images are relatively broad they can easily be seen at low imaging magnifications and large areas of the specimen can be rapidly examined. In an important extension of this technique the defects can be further characterized by the technique of deep level transient spectroscopy (DLTS) in

which the electron beam is used as a source of charge carriers to populate all of the trapping levels of the semiconductor. A measurement of the transient current which flows when the beam is switched off, due to thermal desorption of carriers from the traps, as a function of the specimen temperature, provides a spectrum which both characterizes different types of traps (electron or hole) and their energy within the band gap [13].

(b) *Cathodoluminescence*

The charge collection mode discussed above is a powerful tool for the examination of semiconductor materials and devices. However, it can only be applied if the material contains p–n junctions or if a Schottky barrier can be fabricated on the sample surface, and for many materials of interest, such as a II–VI compound semiconductor, this is not readily possible. In these cases, and also for the study of geological and organic materials, the technique of cathodoluminescence (CL), in which the light emitted from a sample is studied, is of value. The luminescence signal is the result of electronic transitions between quantum mechanical energy states separated by a few electron volts, and two basic types can be distinguished: intrinsic or edge emissions, and extrinsic or characteristic luminescence.

Intrinsic luminescence is associated with the recombination of electrons and holes across the gap and occurs as a band of excitation with its intensity peak at a photon energy

$$h\nu \approx E_{\text{gap}} \tag{4}$$

where E_{gap} is the band-gap energy of the material. In direct gap semiconductors (such as InP, GaAs, CdS) this transition usually results in radiation being produced, but in indirect gap semiconductors (such as Si, Ge, GaP) the simultaneous emission of a photon and a phonon is required to conserve momentum in the transition and the probability of this process is small so the emission is relatively weak.

Extrinsic luminescence depends on the presence of impurities and can occur in both direct and indirect semiconductors. The emission bands in these cases are activated by the presence of impurity atoms or other defects and the luminescence is characteristic of the atom with which it is associated. Extrinsic radiation is much more intense than intrinsic radiation and varies about linearly with the concentration of impurity atoms present. A phosphor is an extreme example of extrinsic luminescence production.

Electron beam excitation leads to emission by both of these modes in all types of semiconductors. The brightness dependence I_{CL} of the cathodoluminescence signal depends on both the beam energy E_0 and the beam current I_b through an expression of the form

$$I_{\text{CL}} = \text{f}(I_b)(E_0 - E_d)^n \tag{5}$$

where the functional dependence on the beam current is about linear, and n is typically between 1 and 2. E_d is the 'dead voltage' of the material, that is, the beam energy below which no luminescence is produced, typically 2 to 5 keV. This 'dead voltage' is the result of the competition between surface recombination and diffusion effects on the electron–hole pairs produced near the sample surface and so is temperature and material sensitive. In general the intensity of the CL signal also

improves as the sample temperature is lowered, especially the weak band-edge emissions. For this reason it is desirable to be able to cool the sample to liquid nitrogen or even liquid helium temperatures. Except for a few exceptional materials the intensity is never very high, so the prerequisites for successful CL operation in the SEM are a highly efficient system for the collection of light, the ability to generate high incident beam currents at moderately high beam energies (10–25 keV), and some provision for sample cooling.

The collection of the CL signal requires care. Although simple systems in which a light guide, or a light sensitive detector, are allowed to look directly at the specimen have often been used these are not reliable because most materials show some fluorescence under electron impact and consequently it is the backscattered electrons, rather than the photons, hitting the collector that generate most of the observed signal. Figure 19 shows schematically a suitable arrangement for performing CL studies in the SEM [23]. The emitted light is collected by an ellipsoidal mirror, with the specimen placed at one focus of the ellipse and the light guide at the other. With this arrangement the light pipe is shielded from the possibility of any impact by incident or scattered electrons, and the solid angle of light collection from the specimen approaches 2π steradians. At high beam energies (1000 keV or more) it may be necessary to use a double mirror system because X-rays generated by scattered electrons on the mirror can also result in the production of spurious CL [24]. If the specimen is to be cooled to cryogenic temperatures then the system is made still more complex by the need to provide radiation shields around the sample, and the need to cool the mirror. Light emerging from the collector can then either be passed to a photomultiplier, for panchromatic imaging, or into a spectrometer for spectral analysis. The luminescence signal is finally detected using photon counting electrons fed either into a scaler or a multichannel analyzer. Although the emission for some common semiconductors is in the wavelength range 300 to 800 nm, for most materials the radiations

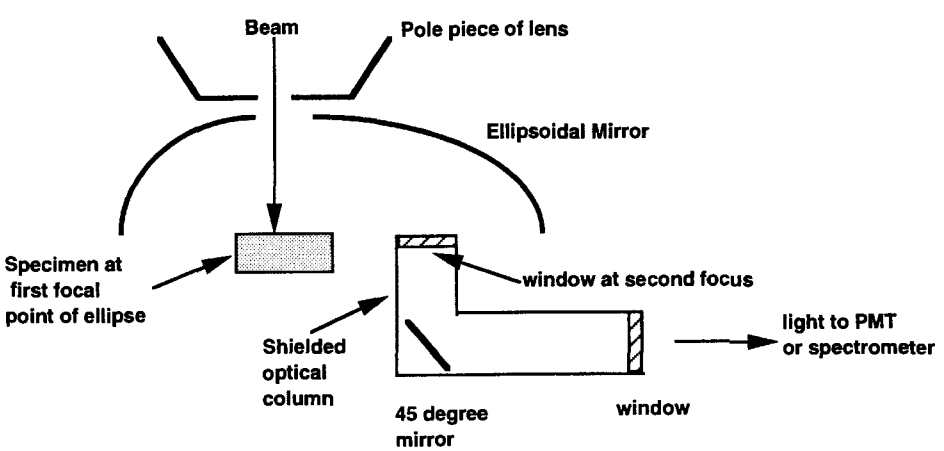


Figure 19. Schematic layout for doing CL in SEM.

of interest occur in the infrared region with wavelengths from 1 to 10 μm or more. It is, therefore, necessary to choose light guides, window materials, and detectors which are suitable for the wavelengths of interest [23].

There are two basic modes of CL operation. In the first, all of the emitted radiation is collected and amplified for display. The CL image produced as the beam scans is therefore a measure of the variation in the total luminescence produced. This mode is simple to set up and is suitable both for the observation of contrast due to extrinsic (dopant produced) CL as well as for intrinsic effects. Figure 20 shows the image of dislocations in a GaAs wafer produced in this way. As for the case of EBIC the dislocation is visible because electron–hole pairs recombine on it instead of producing photons by radiative recombination, the defect therefore appears darker than the surrounding perfect material. The spatial resolution of this kind of image is limited by the diffusion of the carriers and by the rather large incident probe size needed to achieve a high enough

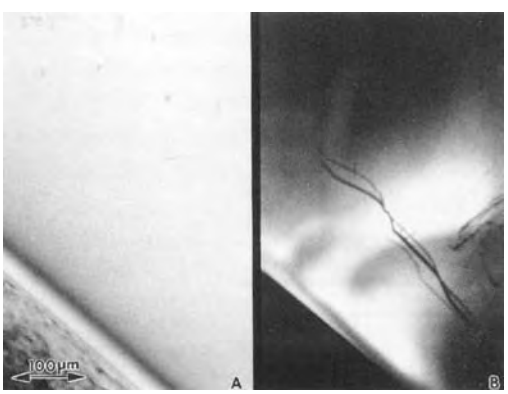


Figure 20. CL image of dislocations in GaAs recorded using panchromatic radiation at liquid nitrogen temperatures in Cambridge S250 SEM.

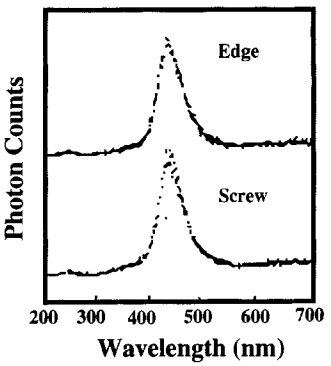


Figure 21. CL spectra from individual dislocations in diamond. Adapted from Yamamoto et al. [24].

beam current (here about 0.1 μA) at the specimen, not by optical diffraction (i.e., the Abbé limit) as would be the case in a conventional optical microscope. A resolution of from 0.1 to 0.5 μm is usually possible.

The other main mode is to collect the spectrum of CL emission from the sample through a grating or prism spectrometer. Figure 21 shows spectra recorded from individual dislocations [24] under the electron beam. Both the wavelength of the peak intensity and the shape of the luminescence peak are seen to vary. These parameters depend on the exact electronic nature of the defect and can therefore be used as a diagnostic. In addition the spectrum allows identification of the characteristic band-edge radiation, and hence a determination of the band-gap of the material, as well as emission peaks due to dopants and impurities. The chemical sensitivity of such an analysis is extremely high, since effects due to dopants below 1 ppb are readily detectable [13]. If sufficient signal is available then imaging can be combined with spectroscopy by forming images from a specified range of wavelengths [24] so allowing the direct

identification of specific features on the specimen.

In summary, cathodoluminescence is a technique which offers many important capabilities for the examination of semiconductors and insulators. The major drawback of the mode has been the complexity of the equipment needed to perform it satisfactorily but the advent of commercial systems may solve this problem and make the technique more accessible.

2.1.5 Conclusions

The scanning electron microscope is a uniquely versatile and powerful tool for the characterization and visualization of materials. It combines high spatial resolution with the ability to look at samples of a practical size, and offers a wide range of imaging modes with which to attack a variety of questions. When the ease of specimen preparation for the SEM and its general user-friendliness are also taken into account the future for this instrument looks promising.

Acknowledgements

Oak Ridge National Laboratory is managed by Martin Marietta Energy Systems Inc. under contract DE-AC05-84OR21400 with the US Department of Energy.

2.1.6 References

[1] M. Von Ardenne, *Z. Phys.* **1938,** *109*, 553.
[2] M. Von Ardenne, *Z. Techn. Phys.* **1938,** *19*, 407.
[3] V. K. Zworykin, J. Hillier, R. L. Snyder, *ASTM Bulletin* **1942,** *117*, 15.
[4] C. W. Oatley, *J. Appl. Phys.* **1982,** *53*, R1.
[5] T. Nagatani, S. Saito, M. Sato, M. Yamada, *Scanning Microscopy* **1987,** *1*, 901.
[6] T. E. Mulvey, C. D. Newman, *Institute of Physics Conference Series* **1973,** *18*, p. 16.
[7] T. E. Everhart, R. F. M. Thornley, *J. Sci. Instrum.* **1960,** *37*, 246.
[8] K. C. A. Smith, *Proceedings of the 5th Annual SEM Symposium* (Ed: O. Johari), IITRI, Chicago **1972,** p. 1.
[9] J. I. Goldstein, D. E. Newbury, P. Echlin, D. C. Joy, A. D. Romig, C. E. Lyman, C. E. Fiori, E. Lifshin, *Scanning Electron Microscopy and X-ray Microanalysis*, Plenum Press, New York **1992,** Chapter 2.
[10] H. Kimura, H. Tamura, *Proceedings of the 9th Annual Symposium on Electron, Ion, and Laser Beams* **1967,** p. 198.
[11] D. C. Joy, *Ultramicroscopy* **1991,** *37*, 216.
[12] D. C. Joy, *Proceedings of the 52nd Annual Meeting MSA* (Ed: G. W. Bailey, A. J. Garratt-Reed), San Francisco Press, San Francisco, **1994,** p. 1032.
[13] D. B. Holt, D. C. Joy, *SEM Microcharacterization of Semiconductors*, Academic Press, London **1989,** Chapters 7 and 9.
[14] K. Koike, K. Hayakawa, *Japan. J. Appl. Phys.* **1984,** *23*, L178.
[15] E. de Harven, D. Soligo, *Am. J. Anat.* **1986,** *175*, 277.
[16] M. D. Ball, M. Wilson, S. Whitmarsh, in *Electron Microscopy and Microanalysis 1987* (Ed: L. M. Brown), Institute of Physics, London **1987,** p. 185.
[17] P. C. Sercel, J. A. Lebens, K. J. Vahala, *Rev. Sci. Instrum.* **1989,** *60*, 3775.
[18] J. Tixier, R. Philibert, *Micron* **1969,** *1*, 174.
[19] D. C. Joy, D. E. Newbury, D. E. Davidson, *J. Appl. Phys.* **1982,** *53*, R81.
[20] D. C. Joy, in *High Resolution Electron Microscopy of Electron Defects* (Ed: R. Sinclair, D. J. Smith, U. Dahmen), Materials Research Society, Warrington, PA, **1990,** p. 199; see also J. T. Czerneska, N. J. Long, E. D. Boyes, P. B. Hirsch, *Phil. Mag. Lett.* **1990,** *62*, 227.
[21] T. E. Everhart, O. C. Wells, C. W. Oatley, *J. Electron. Control* **1959,** *7*, 97.
[22] H. J. Leamy, *J. Appl. Phys.* **1982,** *53*, R51.
[23] B. G. Yacobi, D. B. Holt, *Cathodoluminescence Microscopy of Inorganic Solids*, Plenum Press, New York **1990,** Chapter 1.
[24] N. Yamamoto, J. C. H. Spence, D. Fathy, *Phil. Mag.* **1984,** *49*, 609.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

2.2 Scanning Transmission Electron Microscopy

2.2.1 Introduction

Although the concept of scanning transmission electron microscopy (STEM) was not new, the possibility of employing the scanning principle for high-resolution imaging of thin specimens in transmission was first realized by Crewe, who introduced the use of field-emission guns for this purpose [1, 2]. Dark-field images, obtained with an annular detector in a STEM instrument, showed the first clear electron microscopy images of individual heavy atoms [3] (Fig. 1). From that time, STEM has developed as an important alternative to conventional, fixed beam transmission electron microscopy (CTEM), with special advantages for many purposes.

In a STEM instrument, a fine electron probe, formed by using a strong objective electron lens to demagnify a small source, is scanned over the specimen in a two-dimensional raster (Fig. 2a). The electron probe is necessarily convergent: the convergence angle is, ideally, inversely proportional to the minimum probe size which determines the microscope resolution. On any plane after the specimen, a convergent beam electron diffraction pattern is formed. Some part of this diffraction pattern is collected in a detector, creating a signal which is displayed on a cathode ray tube screen to form the image using a raster scan matched to that which deflects the incident electron beam. If the detector samples the directly transmitted beam (i.e., if it comes within the central spot of the diffraction pattern) a bright-field image is formed. Detection of any part, or all, of the electrons scattered outside the central beam gives a dark-field image.

The use of a field emission gun (FEG) for high resolution STEM is necessary in order to provide sufficient signal strength for viewing or recording images in a convenient time period. The effective electron source of a FEG has a diameter of 4–5 nm as compared to 1–5 μm for guns with LaB_6 or tungsten hair-pin filaments. The amount of demagnification needed to form the electron probe is not large. Because the FEG source has a brightness which is a factor of 10^4 or 10^3 greater than that of a W hair-pin filament or a LaB_6 pointed filament, the total current in the electron beam is greater when beam diameters of less than about 10 nm are produced. The current in a beam of 1 nm diameter is typically about 0.5 nA. With a FEG, bright- or dark-field STEM images can be recorded in a few seconds or even at TV rates (i.e., 30 frames per second).

As suggested by Fig. 2b, the essential components of a STEM imaging system are the same as for a CTEM instrument,

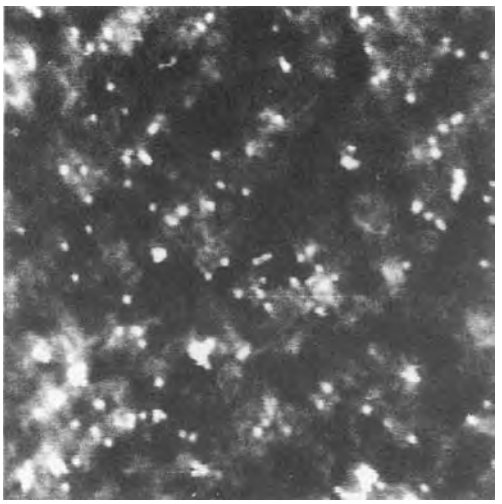


Figure 1. Annular dark-field image of uranium atoms on an amorphous carbon film. Full width of image: 31.5 nm. (Courtesy of J. Wall, Brookhaven National Lab.).

but with the electrons traveling in the opposite direction. In this diagram condenser and projector lenses have been omitted and only the essential objective lens, which determines the imaging characteristics, is included. The STEM detector replaces the CTEM electron source. The STEM gun is placed in the detector plane of the CTEM, and the scanning system effectively translates the STEM source to cover the CTEM recording plate.

It was pointed out by Cowley [4] (see also Zeitler and Thompson [5]) that application of the reciprocity principle implies that, for the same lenses, apertures and system dimensions, the image contrast in STEM must be the same as for CTEM. The reciprocity principle applies strictly to point emitters and point detectors. It states that, for any essentially scalar system, the wave amplitude at a point B due to a point source at A is identical to the wave amplitude at A due to a point source at B. For an incoherent source of finite diameter and a finite incoherent detector, the reciprocity principle may be considered to apply to each point of the source and each point of

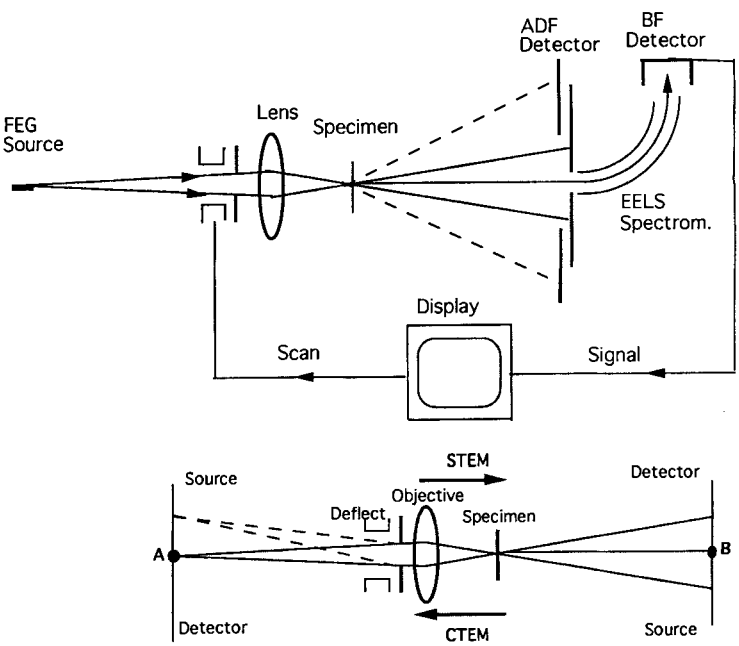


Figure 2. (a) Diagram of the essential components of a STEM instrument.
(b) Diagram suggesting the reciprocity relationship between STEM (electrons going from left to right) and CTEM (electrons going from right to left).

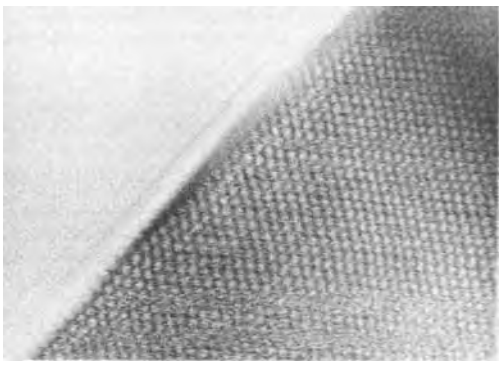


Figure 3. Bright-field STEM image of a small MgO crystal in [1 1 0] orientation [95].

the detector separately. Thus the effect on the image intensity of increasing the STEM detector size is the same as that of increasing the CTEM source size, and so on. For a particular STEM configuration the image contrast can often be inferred by finding the equivalent CTEM geometry, for which the imaging theory has most probably been well established. Thus for bright-field STEM imaging with a small detector, the image is the same as for BF CTEM with the same small angle of convergence of the incident beam (Fig. 3). A large source size for STEM would degrade the image resolution in the same way as a CTEM detector having poor resolving power.

Practical experimental considerations, however, lead to clear advantages or disadvantages of the STEM instruments, relative to CTEM for some imaging modes. For example, dark-field images can be obtained with high collection efficiency in STEM by collecting all the electrons scattered outside of the incident beam spot of the diffraction pattern. The equivalent CTEM configuration would require an incident beam coming from all directions outside the cone of the objective aperture, which is difficult to realize and very inefficient in its use of the incident electrons.

Important differences in the fields of application of STEM and CTEM arise from the different form taken by the image signal. For CTEM a two-dimensional detector such as a photographic plate is used to record intensities at all image points in parallel. In STEM the image information is produced in serial form as a time-dependent voltage or current variation. For many years this gave STEM the unique possibility of online image processing to manipulate the image contrast for special purposes. Now the use of CCD detectors allows a serial read-out and online image processing for CTEM also: however, for STEM further possibilities exist. Thus in STEM several detectors may be used simultaneously to produce signals which may be added, subtracted, multiplied or otherwise manipulated.

A variety of STEM signals may be obtained in addition to the bright-field or dark-field signals derived from the elastic scattering of electrons in the specimen. STEM instruments are normally fitted with an energy-loss spectrometer which not only allows microanalysis of very small specimen regions by electron energy-loss spectrometry (EELS), but also allows images to be formed with electrons that have lost particular amounts of energy, characteristic of particular elements or of particular electronic excitations. Also images may be formed by detecting secondary radiations, such as low-energy secondary electrons, Auger electrons, or characteristic X-rays, as will be discussed below. The serial nature of all these image signals provides possibilities

for quantitative comparisons and correlations of information on specimen compositions, crystallography and morphology which can be very valuable.

The serial mode of imaging in STEM has some practical disadvantages. The recording times for images are usually longer than for CTEM: up to 20 s. The image quality can be degraded by fluctuations in the emission from the field-emission tip, resulting from fluctuations of the work function of the tip surface as molecules of residual gas are absorbed or desorbed. Such fluctuations may give the images a streaky appearance. Also there may be a steady decay of emission current so that the correlation of signals over time is difficult. These effects may be minimized if the gun is operated in very high vacuum, but it is difficult to eliminate them entirely. On the other hand, specimen drift can have an adverse effect in CTEM, smearing out the image and degrading the resolution, whereas in STEM, if the image is recorded with a single slow scan of the image field, the effect of a specimen drift is to produce only a slight distortion of the image.

2.2.2 Scanning Transmission Electron Microscopy Imaging Modes

Bright-field STEM images from thin specimens, obtained with a small axial detector, show the same contrast features as for CTEM, as expected from the reciprocity relationship. Phase-contrast effects including the reversal of contrast with defocus and Fresnel fringes, and amplitude contrast due to diffraction and inelastic scattering, are similar (Fig. 3). If the detector aperture is very small, to simulate the case of parallel-beam CTEM, the image signal is generally too weak and noisy for convenience. Hence, there is a tendency to use larger detector angles with the result that the image contrast is somewhat reduced (although it can be restored by online image processing) and the resolution is slightly improved, although not so readily calculated (see next Section).

For STEM, an EELS detector is normally present so that the BF image may be formed with all electrons transmitted through the specimen or else with only those electrons which have lost less than about 1 eV in energy. For thin specimens, the difference in these two cases is small. For specimens of thickness comparable with the mean free path for inelastic scattering (of the order of 100 nm for 100 keV electrons), the difference becomes significant. The resolution and contrast of the images are degraded for thick specimens by two factors.

First, there is a loss of resolution from the geometric effect of multiple scattering: the point of origin of the scattered electron becomes indefinite because after a second scattering process, for example, it cannot be determined whether the electron comes from the first scattering point or from the second. The different scattering processes appear to occur at different lateral positions. This effect is the same for STEM and CTEM. Second, the inelastic scattering produces a spread of energies of the transmitted electrons. For a 100 nm thickness for 100 keV electrons, for example, the average energy loss is about 30 eV and there is an average spread of energies of about the same magnitude. For CTEM, the chromatic aberration of the objective

lens then degrades the resolution, but in STEM, because the objective lens comes before the specimen, the effect of such an energy spread is negligible. Hence, in general, the resolution and/or contrast is better for STEM than for CTEM for a given specimen thickness: or, for a given resolution level, the thickness of specimens that can be used is greater for STEM than for CTEM [6, 7]. Because the chromatic aberration effect decreases with accelerating voltage, being proportional to the ratio of the energy spread and the incident beam energy, the advantage of STEM decreases, but even so, a STEM operating at 0.5 MeV has the same 'penetration' as a CTEM at 1 MeV [8].

In the original work of Crewe et al. [3], the detector used was an annular one designed to collect all electrons scattered out of the incident beam. Simplifying approximations which are good for isolated heavy atoms, but may break down for groups of atoms [9], suggest that the image intensity is then proportional to $Z^{3/2}$, where Z is the atomic number of the elements present and the image resolution should be better than for bright-field by a factor of 1.4 or 1.5. The efficiency of collection of the dark-field signal is much better than for dark-field CTEM, for which only a small part of the diffraction pattern can be collected by the objective aperture and used to form the image.

The Z-dependence of the annular dark-field (ADF) image suggested its use for detecting heavy-atom particles in a matrix of light-atom material such as occur, for example, in supported metal catalysts. However, if the light-atom material is microcrystalline, as is often the case, the variations of diffraction intensities produce large fluctuations in the ADF image intensity, obscuring the heavy atoms. Howie [10] suggested that this effect could be avoided if the inside hole of the annular detector is made so large that the lower-angle region, where the diffraction spots occur, is not included (Fig. 4). When a high-angle annular dark-field (HAADF) detector is used, good Z-dependent contrast is obtained, independent of crystallinity [11] (Fig. 5).

The signal collected comes, in part, from the high-angle elastic scattering which has an average value depending on the square of the atomic scattering factor, $f(u)$, where u is the angular variable, equal

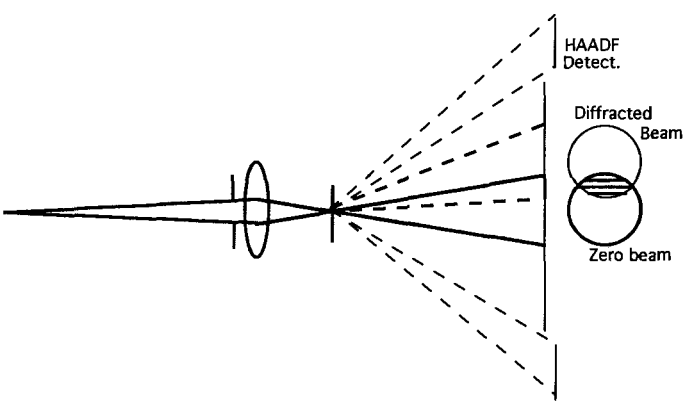


Figure 4. Diagram of a STEM system suggesting the formation of overlapping diffracted beam disks, with interference fringes in the area of overlap, and the collection of high-angle scattered radiation to form HAADF images.

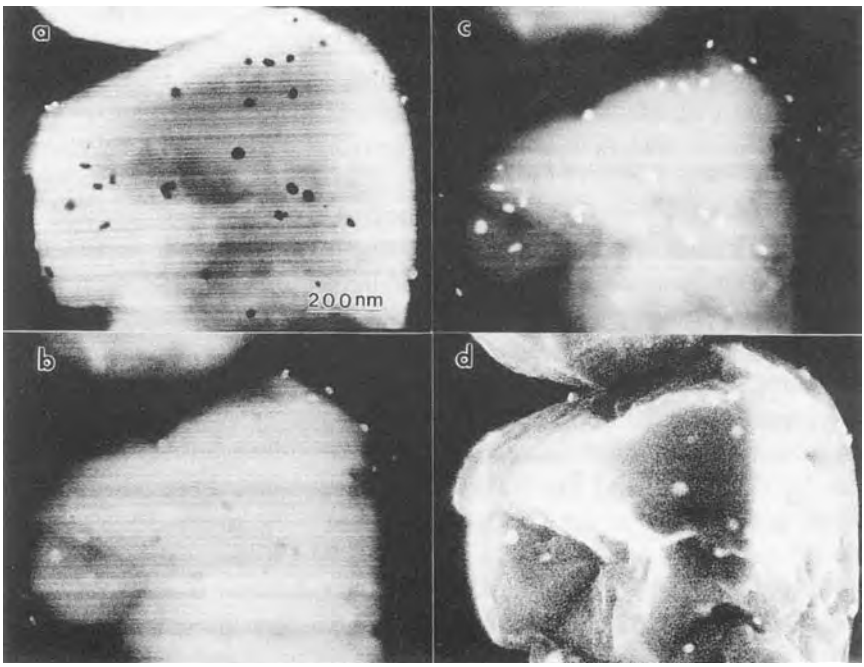


Figure 5. (a)–(c) ADF STEM images of Pt particles on a thick NaY–zeolite crystal with inner collector angle 20 mrad for (a) and increasing for (b) and (c), plus a SEM image (d), of the same specimen showing the Pt particles on one surface. Note the reversals of contrast of the Pt particles as the thickness and detection angles change. (Courtesy of J. Lui [96]).

to $(2/\lambda)\sin(\phi/2)$ for a scattering angle ϕ. A predominant part of the signal, however, comes from the thermal diffuse scattering which is intrinsically a high-angle scattering because, even for first-order scattering, the intensity depends on the square of $u \cdot f(u)$ [12, 13]. More recently it has been shown that the HAADF mode can give good high-resolution imaging of crystals in principal orientations in which differences in Z of the columns of atoms parallel to the beam can be clearly indicated [14, 15].

It is a virtue of STEM that the form of the detector may be varied readily to give special types of information in the images. In Fig. 5 it is seen that, for a relatively thick specimen, the influence of multiple scattering to higher angles can lead to striking changes of contrast as the inner diameter of the ADF detector is changed. For thin specimens, if a thin annular detector is used, with only about 10% difference between inside and outside diameters, images may be obtained from different ranges of diffraction angles so that components of the specimen which diffract differently may be distinguished (Fig. 6).

Circular detectors in the bright-field area, split into two semicircular halves, were proposed by Dekkers and de Lang [16]. The signals from the two halves may be added or subtracted. A simple, geometrical optics description serves to give a useful approximation to the image intensities. A linear increase or decrease in the projected potential in the specimen acts

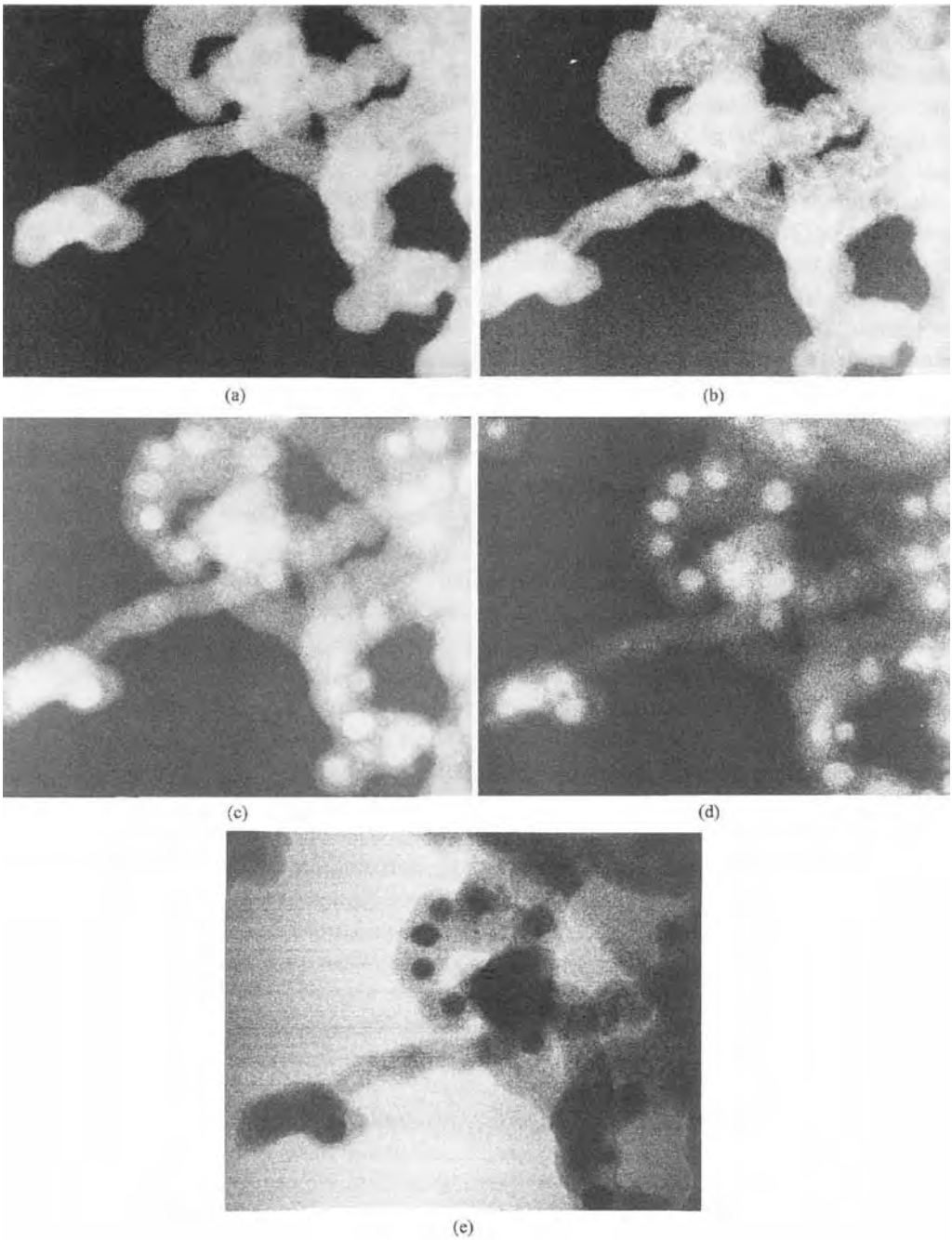


Figure 6. STEM images of a specimen of carbon particles formed in a high-pressure arc with Mn present. (a)–(c), Images with a thin annular detector for average d-spacings of 0.6, 0.3 and 0.1 nm, showing, respectively, the amorphous carbon, the graphitic carbon and the Mn carbide particles. (d) HAADF image, and (e) bright-field image.

like a prism for electrons, deflecting the incident beam one way or the other. If the deflection has some component in the direction perpendicular to the slit between the two semicircular detectors, the signal in one detector is increased and the signal in the other is decreased. The difference signal then corresponds to the gradient of the projected potential of the specimen and is either positive or negative. The sum of the signals, proportional to the projected potential for a weak phase object, can be used as a reference signal.

This differential imaging mode is useful in some cases for detecting particles in a matrix since it gives sharply defined positive or negative signals at the edges of the particle. It can also be used to detect magnetic fields in a specimen which deflect the incident beam. Chapman et al. [17] have developed the method, called differential phase contrast imaging, into a highly effective means for studying the magnetic fields and their changes in the domain structures of thin ferromagnetic films. By using a circular detector split into four quadrants, and adding or subtracting the signals from the various quadrants, they could define the directions and strengths of the magnetic fields. One difficulty with this approach is that the signal from the magnetic field variation may be confused by signals from the variation of projected potential arising from variations of specimen thickness or structure. A further subdivision of the detector into eight regions, with two concentric sets of four quadrants, allows the separation of the magnetic signal, which occurs mostly at lower angles of scattering, from the structural signal [18]. Also, even more complicated multiple detectors have been proposed [19].

2.2.3 Scanning Transmission Electron Microscopy Theory

Following Cowley [20], we may write relatively simple expressions to derive the form of the image contrast for various STEM detectors if we confine our treatment to thin objects for which the effect of the specimen is well represented by a transmission function, $q(x, y)$, which multiplies the incident electron wave-function. The electrostatic potential within the specimen modifies the energy and wavelength of the electrons, acting like a refractive index slightly greater than unity, and so modifies the phase of the electron wave by an amount proportional to the projection of the potential distribution in the incident beam direction

$$\int \phi(x, y, z)\, \mathrm{d}z \equiv \phi(x, y)$$

Then the transmission function is

$$q(x, y) = \exp[-\mathrm{i}\sigma\phi(x, y)]$$

where σ is the interaction constant, equal to $2me\lambda/h^2$. This is the so-called phase-object approximation, valid only for very thin specimens but without the limitation of the weak-phase object approximation which assumes that $\sigma\phi(x, y) \ll 1$ and is valid for only light atom materials. In the following, for convenience, we deal with only one-dimensional functions. The extension to two dimensions is obvious.

For the very small effective source size of a field emission gun, the convergent beam incident on the specimen may usually be assumed to be completely coherent, as if coming from a point source, represented by a delta function. The coherence function at the objective aperture

position is given, according to the Zernike–van Cittart theorem, by the Fourier transform of the intensity distribution of the effective source. For the usual dimensions of an FEG illumination system, the width of the coherence function is of the order of several millimeters and so is much greater than the diameter of the objective apertures used for optimum STEM imaging. For the very much larger source sizes common for thermal electron sources, the width of the coherence function at the objective aperture is normally only about 1 μm, so that it is a good assumption that electron waves arriving at the specimen from all parts of a 50 or 100 μm aperture are completely incoherent.

If the electron source is assumed to be a delta function, the electron wave coming through the objective lens is given by the transfer function of the lens, $T(u)$

$$T(u) = A(u)\exp[i\chi(u)] \tag{1}$$

where $u = (2/\lambda)\sin(\phi/2)$ for scattering angle ϕ and the aperture function $A(u) = 1$ if $u < u_0$ and 0 if $u > u_0$, and the phase factor,

$$\chi(u) = \pi\Delta\lambda u^2 + \frac{\pi}{2}C_s\lambda^3 u^4$$

where Δ is the defocus (negative for underfocus, i.e., for a weakening of the objective lens) and C_s is the spherical aberration constant. In the approximation normally used for electron microscopy at the current levels of resolution, the other higher-order or off-axis aberrations, which would add further terms in the phase function, are ignored and it has been assumed that the astigmatism has been corrected. Then the wave amplitude incident on the specimen is given by Fourier transform of $T(u)$ as the spread function $t(x)$ which represents the smearing-out of the ideal image of the source due to the defocus and aberrations of the lens. Translating the incident beam by an amount X, the wave transmitted through the specimen is $q(x)\,t(x - X)$. Fourier transforming and squaring this function gives the intensity distribution on the detector plane as

$$I_X(u) = |Q(u) * T(u)\exp(2\pi i u X)|^2 \tag{2}$$

Here the $*$ denotes the convolution operation defined by

$$f(x) * g(x) \equiv \int f(X)\,g(x - X)\,dX$$

If the detector has a transmission function, $D(u)$, the signal detected as a function of the incident beam position is

$$S(X) = \int D(u) I_X(u)\,du \tag{3}$$

It is useful to distinguish the transmitted beam from the scattered waves, so we put $q(x) = 1 - p(x)$, of which the Fourier transform is $Q(u) = \delta(u) - P(u)$. Substituting in Eq. (3) then gives

$$\begin{aligned} S(X) = &\int D(u)\,A(u)\,du \\ &+ \int D(u)|P(u) * T(u) \\ &\times \exp(2\pi i u X)|^2\,du \\ &- \int D(u)\,T^*(u)\exp(-2\pi i u X) \\ &\times [P(u) * T(u)\exp(2\pi i u X)]\,du \\ &- \text{c.c.} \end{aligned} \tag{4}$$

where c.c. indicates the complex conjugate of the previous term.

If one makes use of the relation

$$Q(u) * T(u) \exp(2\pi i u X) = \exp(2\pi i u X)[q(X) * t(X) \times \exp(-2\pi i u X)]$$

Eq. 4 can be written as

$$S(X) = \int D(u)\, A(u)\, \mathrm{d}u + \int D(u)|p(X) * t(X) \times \exp(-2\pi i u X)|^2\, \mathrm{d}u - p(X) * \left[t(X) \int D(u)\, T^*(u) \times \exp(-2\pi i u X)\, \mathrm{d}u\right] - \text{c.c.} \quad (5)$$

In the expressions on the right hand side of Eqs. (4) and (5), the first term comes from the incident beam. The second term comes from the scattered waves, both inside and outside the incident beam cone. It is a term of second order in $p(X)$ and so represents an integration over intensities. The third term and its complex conjugate are of first order in $p(X)$, derived from the coherent interference effect of the superimposed incident and scattered waves and so contribute to the intensity only within the central beam spot defined by $A(u)$.

From these expressions, it is straightforward to derive the image intensities obtained for various forms of the detector function, $D(u)$. For example, for a very small detector located on the axis, a good approximation is $D(u) = \delta(u)$. Then Eq. (5) reduces to

$$S(X) = |q(X) * t(X)|^2 \quad (6)$$

which is identical with the expression for bright-field CTEM with a parallel beam illumination. The specimen transmission function is smeared-out by the spread function and its modulus is squared. In the weak phase object approximation, the intensity of the image is given simply as $I(X) = 1 + 2\sigma\phi(X) * s(X)$, where $s(X)$ is the Fourier transform of $\sin\chi(u)$, the imaginary part of the transfer function, so a direct smeared-out representation is given of the projected potential.

For an annular detector which collects all the electrons scattered outside the central beam spot, the product $D(u)\, T(u)$ is zero so that only the second term remains in Eqs. (4) and (5). In general, this term is not easily evaluated, but if the approximation is made that the contribution of scattered intensity from within the central beam spot is proportional to the signal from outside the central beam spot (as is approximately the case for scattering from a single isolated atom) then one can assume that $D(u)$ is a constant and

$$S(X) = \int |P(u) * T(u) \exp(2\pi i u X)|^2 = \int |p(X)\, t(x - X)|^2\, \mathrm{d}X = |p(X)|^2 * |t(X)|^2 \quad (7)$$

where the second equality follows from Parseval's theorem and the result implies that the square of the scattering function $p(X)$ is imaged with a spread function equal to the intensity distribution of the incident beam. The annular dark field (ADF) imaging has hence been described as 'incoherent' imaging. For the weak phase-object approximation, $\sigma\phi(X) \ll 1$, the image intensity depends on $\phi^2(X)$, so that a small bright dot should appear in the image for a positive peak in potential

(e.g., an isolated atom) or for a negative peak (e.g., a hole in thin perfect crystal). On the assumption that $t(X)$ is a Gaussian function, the resolution for the ADF image is seen to be better than for the bright-field image by a factor of $\sqrt{2}$.

The assumption that $D(u)$ may be replaced by a constant is a reasonable one when the scattering from the object gives an intensity distribution which falls off smoothly from the center of the diffraction pattern as is the case for single isolated atoms or for amorphous materials. It may fail significantly, however, for particular cases such as two closely-spaced atoms, for which the diffraction pattern has sinusoidal oscillations of periodicity comparable with the dimensions of the aperture [9].

If $D(u)$ represents an axial, circular hole, smaller than the central beam disc, the intensity distribution in the image can be evaluated by computer calculations [21]. Some indication of the effect can be seen from the form of the last term in Eq. (5). If the integral gives a function which falls away from a central peak, multiplying $t(X)$ by this function has the effect of making the spread function narrower and hence of improving the resolution. However, it turns out that as the detector hole diameter is increased, the constant decreases for weak phase objects and becomes zero for a detector diameter equal to that of $A(u)$. On the other hand, the second term of Eq. (4) or Eq. (5), for $D(u) = A(u)$, becomes equal to the negative of Eq. (7), so that a bright-field image is given with the same resolution as the dark-field image although with relatively poor contrast because of the constant background intensity from the first term of Eq. (4) or Eq. (5) [22].

To improve the bright-field resolution even further, and to make use of the first-order terms in the intensity expressions, it is necessary to find a detector function, $D(u)$, such that the integral in the third term of Eq. (5) represents a peaked function much sharper than $t(X)$. It was shown [20], for example, that if $D(u)$ represents a thin annulus, with inside and outside diameters differing by only about 10% and an average radius equal to the optimum bright-field aperture size, the bright-field resolution should be improved by a factor of 1.7. This thin-annular bright-field STEM mode is equivalent, by reciprocity, to the CTEM bright-field mode using a hollow-cone incident beam, which has been explored extensively [23, 24].

For thicker crystals (thickness greater than 2–5 nm for 100 keV electrons) a simple transmission function, such as the phase-object approximation, cannot be used. As in CTEM, the diffraction pattern and image intensities must be calculated using one of the approximations to the three-dimensional, many-beam dynamical diffraction theory such as the Bloch-wave or multislice methods [25, 26]. For small detector aperture sizes it is convenient to make use of the reciprocity relationship and make the calculations for the equivalent CTEM configuration with a nearly parallel incident beam.

For imaging with large or more complicated detector configurations or for the convergent-beam diffraction patterns, one approach is to make dynamical calculations for each incident beam orientation and then add the diffraction amplitudes or intensities (depending on the assumptions of coherence of the incident beam) for each diffracted beam direction. In an alternative

approach, a multislice calculation is made for an incident beam amplitude described by the spread function, $t(x, y)$. Since $t(x, y)$ is nonperiodic, it is placed within an artificially large unit cell corresponding to a large number of the crystal lattice unit cells, making use of the assumption of periodic continuation [27]. The effect of imposing this large periodicity on the structure in real space is to sample the continuous scattering function in reciprocal space at a finely spaced lattice of points. Hence, the number of effective 'diffracted beams' becomes very large, although not too large to be handled conveniently by modern computers.

For the relatively simple case of lattice fringe imaging, we may consider the case of a periodic structure, periodicity a, giving a one-dimensional row of diffraction spots with an objective aperture size such that the diffraction spot discs overlap. The diffraction pattern amplitude is then described by

$$\Psi(u) = T(u)\exp(2\pi i u X) * \sum_h F_h\,\delta(u - h/a) \tag{8}$$

From this may be derived the intensity distribution which, as will be discussed later, includes all the coherent interactions of the diffracted beams with the incident beam and with each other. In the region of overlap of diffraction spot disks, patterns of interference fringes appear (Fig. 4). For the moment, however, we confine our attention to the midpoints of the regions of overlap. The midpoint of the region of overlap of the disks due to the h and $h+1$ reflections comes at $u = +\frac{1}{2}a$ relative to the h reflection and at $u = -\frac{1}{2}a$ relative to the $h+1$ reflection, so that the intensity at that point is

$$\begin{aligned}|\Psi(u+\tfrac{1}{2}|^2 &= |F_h \exp[i\chi(+\tfrac{1}{2}a) + \pi i X/a] \\ &\quad + F_{h+1}\exp[i\chi(-\tfrac{1}{2}a) - \pi i X/a]|^2 \\ &= |F_h|^2 + |F_{h+1}|^2 + 2|F_h| \\ &\quad \times |F_{h+1}|\cos(2\pi X/a + \alpha)\end{aligned} \tag{9}$$

where α is the phase difference between the two reflections, since $\chi(u)$ is symmetrical. Hence, if a small detector is placed at the midpoint of the overlap region, the STEM image shows sinusoidal fringes of period a and a shift of the fringes relative to an origin point which indicates the relative phase of the diffraction spots [28]. It has been proposed that, on this basis, it may be possible to derive the relative phases of all diffraction spots in a diffraction pattern. In the case of the weak phase-object approximation, this information would allow the summing of the Fourier series with coefficients F_h to give the projection, $\phi(x, y)$, of the crystal potential. In this way one could make an unambiguous structure analysis of a thin crystal, avoiding the 'phase problem' which hinders the structure analysis of crystals based on kinematical X-ray or electron diffraction data [29, 30]. The practical difficulties which have prevented the realization of this scheme include the difficulty of obtaining sufficiently thin single crystals of nontrivial structure which are sufficiently resistant to radiation damage by the incident beam.

2.2.4 Inelastic Scattering and Secondary Radiations

In addition to the elastic scattering, the beam in a STEM instrument undergoes a

number of inelastic scattering processes. Thermal diffuse scattering with energy losses of about 0.02 eV, produces a broad diffuse background to the electron diffraction pattern extending to high angles of typically 10^{-1} radians, with peaks of intensity around the Bragg diffraction spots. In the bulk of conducting samples, collective electron oscillations, or plasmons, are generated resulting in energy losses in the range of 5 to 30 eV and an angular distribution of the energy loss electrons of 10^{-3} radians or less. At the specimen surfaces, surface plasmons are generated with somewhat smaller energy losses and comparable scattering angles. For nonconducting or conducting specimens, electrons from the outer shells of atoms, or from the conduction bands of solids, may be excited with the same range of energy losses of 5 to 30 eV and a somewhat broader distribution of scattering angles which is still much narrower than for elastic scattering.

The excitation of inner shell electrons of the specimen atoms gives energy-loss peaks in the range of about 50 to several thousand eV with energies depending on the atomic number and the electron shell (K, L, M, ...) involved. The dependence of the inner-shell energy losses and of the energies of the emitted X-rays on the nature of the atoms present provide the basis for microchemical analysis techniques.

Spectroscopy of the transmitted electrons using electron energy-loss spectroscopy (EELS) is performed by passing the electrons transmitted through the specimen through a magnetic quadrant spectrometer. The X-rays emitted from the specimen are detected by placing the small energy-sensitive detector used for energy dispersive spectrometry (EDS) as close to the specimen as possible. These microanalytical methods, as applied to the thin samples in CTEM or STEM instruments, are described in Secs. 1. and 2.4 of this Chapter. Also the signal from electrons having a particular energy loss, or from the X-rays of a characteristic wavelength, may be used to form an image which displays the distribution of a particular element in the sample.

The particular virtue of the STEM instruments for these purposes arises because, with a cold field-emission gun giving a high intensity of electrons concentrated into a very small probe, the microanalysis may be performed for specimen regions of diameter as small as 1 nm or less and the images showing distributions of the elements may be produced in a convenient time without excessive noise. Although the total electron current from an FEG source is much less than from a thermal source, the much greater brightness of the FEG source ensures that the electron current that can be concentrated in a beam of small diameter is greater for beam diameters of less than a few tens of nanometers. Thus STEM instruments are much more effective if the microanalysis is to be carried out, or atom-specific imaging is to be performed, with high spatial resolution. Discussion and example of analysis and imaging with high spatial resolution have been given for EELS [31, 32] and for X-rays [33].

The signal strengths for energy losses due to plasmons and outer-shell excitations are much higher than for inner shells, but the energy loss distributions are in general rather diffuse and not often clearly characteristic of particular materials. A few materials such as elemental Al and Si

give sharp plasmon peaks which have been used for discrimination in special circumstances [34]. Also the relationship of the energy loss spectra to the electronic states of solids and their surfaces has been explored [35]. The use of very fine beams in a STEM instrument has suggested the possibility of exploring the energy losses, and hence the nature of the excitations of specimen surfaces, by passing the beam close to the edge of a particle or parallel to the face of a small crystal [36, 37]. Such observations have been compared with theoretical models of surface potential distributions and their interactions with high-energy beams [38].

Apart from the characteristic X-rays, the secondary emissions from materials struck by high-energy electrons include low-energy secondary electrons, which provide the basis for SEM imaging (see Sec. 2.1 of this Chapter) and Auger electrons which may be detected to provide scanning Auger electron microscopy (SAM). Both the secondary electrons and the Auger electrons which are detected, being of relatively low energy, arise from atoms close to the specimen surfaces. The secondary electron images give information on surface morphology and are dependent, to a limited extent, on the composition and structure of the near-surface layers of atoms (see Fig. 5). The Auger electrons, which are derived from inner-shell excitations, have energies which are characteristic of the atomic species and so provide information on the compositions of the surface layers. When applied in microscopes having an ultrahigh vacuum specimen environment and the means for providing atomically clean surfaces, SEM and SAM become powerful tools for surface science [39].

Modern SEM instruments, equipped with field-emission guns and operating usually at about 15 to 30 keV, have shown surface morphology with a spatial resolution of 1 nm or better [40]. In a 100 keV instrument, resolutions of 0.6 nm have been demonstrated [41]. There is thus very little basis for the argument that, because the excitation processes leading to secondary electron emission are delocalized, the SEM resolution must be limited to about 10 nm or more. An explanation for the high resolution attainable in SEM with a FEG source has been provided by the electron coincidence measurements of Scheinfein et al. [42]. With a suitable configuration of detectors in a STEM instrument it is possible to obtain SEM images from either the top or the bottom surface of a specimen (or both) in parallel with normal bright-field or dark-field transmission images [43] (Fig. 13).

Since energy spectroscopy of Auger electrons (AES) and SAM imaging allow determinations of the chemical compositions and atomic distributions of the few top layers of atoms on a specimen surface, they can be valuable for ensuring the cleanliness of surfaces with a sensitivity of a fraction of a monolayer of impurity and for following the details of surface chemical reactions. The conventional methods of exciting and detecting Auger electrons using cylindrical or hemispherical electron energy analyzers, are limited by the noisiness of the signals to spatial resolutions of 10 nm or greater. In a UHV STEM instruments, the high-brightness source and the use of special high-efficiency electron detectors have allowed spatial resolutions of 1 nm or better to be achieved [44].

2.2.5 Convergent-Beam and Nanodiffraction

The diffraction patterns that form on the detection plane of a STEM instrument are necessarily convergent beam electron diffraction (CBED) patterns with diffraction spot disks of diameter determined by the objective aperture size. With a two-dimensional detector such as a phosphor or CCD system, these patterns can be observed and recorded and used, in the same way as CBED patterns obtained in a CTEM instrument, for accurate determination of crystal lattice constants or crystal thickness or for absolute determinations of space-group symmetries (see Sec. 1.5 of this Chapter). The use of a field-emission gun, rather than a thermal emission gun, introduces two new aspects. First, the region from which the CBED pattern is obtained may be much smaller; 1 nm or less in diameter rather than 10 nm or more. Second, the FEG source gives an essentially coherent incident convergent beam, rather than an essentially incoherent one, so that the effects of coherent interference may be seen in the patterns.

For an ideally perfect thin crystal, there is no difference in the diffraction patterns for coherent and incoherent radiation if the diffraction spot disks do not overlap. If they do overlap, interference fringes appear in the region of overlap if the incident beam is coherent [45, 46]. The coherent interference effects in diffraction patterns will be considered in detail in the next part of this Section. For now we concentrate on the possibilities for obtaining diffraction patterns from regions 1 nm or less in diameter (i.e., on nanodiffraction).

Such patterns can be obtained from any selected region of a STEM image merely by stopping the scan at the designated spot, or else series of patterns can be recorded at up to TV rates (i.e., $30\,s^{-1}$), as the electron beam is scanned slowly across the specimen, by use of a low-light-level TV camera and a video cassette recorder (VCR) [47]. Diffraction patterns such as those of Fig. 7 were obtained with a nominally 1 nm beam; that is, with a beam convergence angle of about 10 mrad so that the beam at the specimen has a half-width at half-height of 0.7 nm and a diameter, measured to the first zero intensity, of 1.6 nm. For a larger objective aperture size the beam diameter at the specimen can be made as small as about 0.2 nm, but then the diffraction spots are so large that they overlap for even small unit cell sizes and it is no longer possible to recognize and measure the spot geometries easily. For large unit cell sizes, a smaller aperture size is preferred.

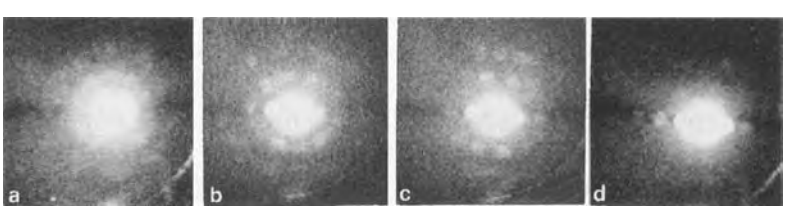


Figure 7. Nanodiffraction patterns obtained as a beam of diameter 1 nm is scanned across a carbon nanotube of diameter approx. 20 nm showing the variation of the structure from the middle of the tube at (a) to one edge of the tube at (d) [92].

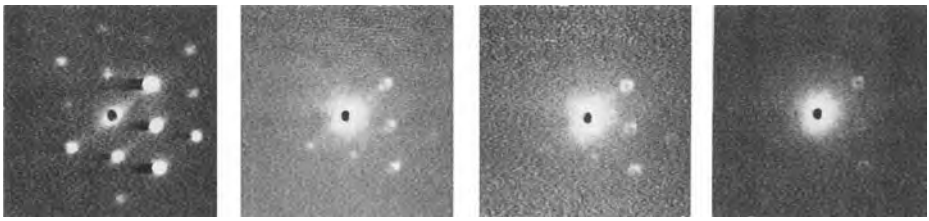


Figure 8. Nanodiffraction patterns obtained as a beam of diameter 1 nm is scanned across the edge of an MgO crystal showing the open circle form of the diffraction spots when the beam covers the crystal edge.

Imperfections in crystals, including crystal edges, faults or other defects, give scattering in between the Bragg diffraction spots of the pattern. Then coherent interference effects, characteristic of the type of defect, are seen. When the region of the specimen illuminated by the incident beam includes the straight edge of a crystal, it is to be expected from kinematical diffraction theory for a parallel incident beam that all diffraction spots are accompanied by streaks perpendicular to the crystal edge. For a coherent convergent incident beam, however, the diffraction disks become hollow circular rings, as in Fig. 8. A crystal fault plane, or an out-of-phase domain boundary in a superlattice structure, gives a similar form for spots corresponding to the particular set of crystal lattice planes for which the defect represents a discontinuity [48, 49]. The spots take on rather more complicated shapes if the beam illuminates part or all of a small crystal of regular shape [50]. The spot shape depends on the wedge angle if the boundary of the crystal is given by two planar surfaces meeting at an edge. Also the lateral extent of the edge is significant.

It follows that nanodiffraction can provide a very effective means for studying very small regions of crystals, such as those containing individual defects, or crystals which have dimensions in the range of 1 to 10 nm, such as, for example, the small metal crystals in supported metal catalyst samples. The nanodiffraction is often coupled with STEM imaging, especially in HAADF mode, which provides a very effective means for locating small heavy-metal particles, and with EELS or EDS microanalysis for determining the particle composition.

2.2.6 Coherent Nanodiffraction, Electron Holography, Ptychology

As the objective aperture size in a STEM instrument is increased, the diffraction spots in the diffraction pattern from a crystal increase in diameter, overlap, interfere and become less distinct, until the periodicity of the crystal is no longer evident. Then the incident beam diameter is much less than the crystal unit cell dimensions, as viewed from the incident beam direction. The pattern may be considered as given only by that grouping of atoms illuminated by the beam. The symmetry and intensity distribution in the pattern change as the beam is moved about within the unit cell [51].

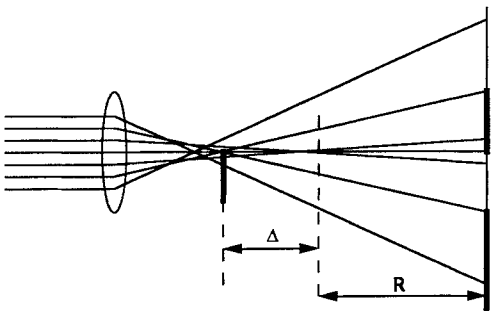


Figure 9. Geometrical optics ray diagram of the shadow-image formation with strong spherical aberration. The infinite-magnification circle is formed when adjacent rays cross over at the specimen plane. Within this circle, the image of the straight-edge is inverted.

For a very large, or no, objective aperture, the diffraction pattern turns into what appears to be a 'shadow image' of the specimen with a magnification, on axis, proportional to the inverse of the defocus value. Off-axis, the magnification is affected by the spherical aberration of the objective lens as suggested by Fig. 9. In over-focus (i.e., with the objective lens current increased) the magnification decreases uniformly from an axial maximum. For under-focus, the magnification increases from the axial value and goes to infinity for some particular angle (depending on the defocus and C_s values), before reversing sign and decreasing with increasing angle. Figure 9 refers only to radial magnification which becomes infinite when two beams which are adjacent in the radial direction have a crossover at the specimen plane. An infinity in magnification in the circumferential direction, given when adjacent beams having the same distance from the axis cross over at the specimen position, occurs at a radius in the image $\sqrt{3}$ times as great. This infinite magnification gives rise to a very obvious infinite-magnification circle, which is of great practical use for determining the defocus and for correcting the lens astigmatism [52].

For thin crystalline specimens, the defocused images show patterns of lattice fringes which becomes distorted by the spherical aberration, astigmatism or other aberrations to give the characteristic bowed shapes of Ronchi fringes (Fig. 10), similar to those familiar from light optics

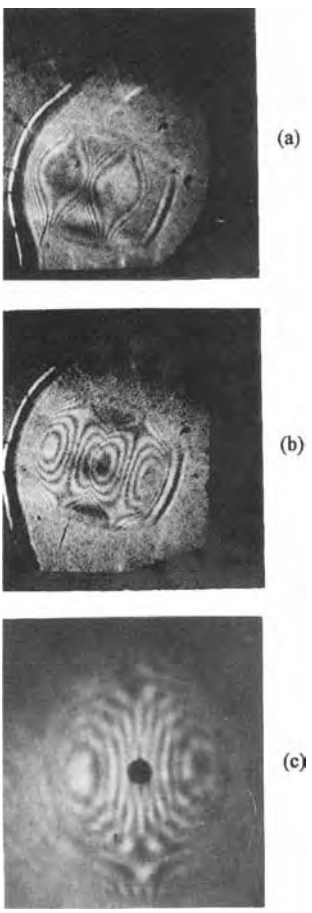


Figure 10. Electron Ronchi fringes formed in the shadow image (or in-line hologram) of a crystal. (a), (b) Ronchi fringes from the (2 0 0) planes of MgO at different defocus values; (c) Ronchi fringes from the 0.8 nm planes of beryl, showing the zero-contrast ellipses [54].

and used for the testing of large telescope mirrors [53]. In STEM instruments, such patterns are highly useful for correcting astigmatism and for determining C_s and defocus values with high accuracy [54, 55].

The shadow images formed with a stationary beam in STEM instruments contain all the coherent interference effects between the transmitted incident beam and the beams scattered by the specimen. They are exactly what Gabor [56, 57] called holograms. Gabor proposed the method of holography as a means for overcoming the limitation of electron microscope resolution due to the irreducible aberrations of the objective lens. He realized that the aberrations produce a perturbation of the phase of the waves interfering to form the image and could possibly be removed if the phases of the waves could be determined by interference with a reference wave of known form. He considered that if a thin object was illuminated by a very small source, the directly transmitted ('unscattered') wave could serve as a reference wave and the effects of its interference with the relatively weak waves scattered by the specimen would be recorded in the hologram.

Gabor proposed a method for reconstructing the original transmission function of the object from the hologram by use of a light-optical system having a built-in spherical aberration to correct for the aberration of the electron-optical lens, thereby enhancing the resolution of electron microscopes. Gabor's suggested technique became feasible only after the development of field-emission guns provided sufficiently small, bright sources of electrons. It was then realized using digitized holograms and computer processing in place of the optical reconstruction methods [58].

The main difficulty associated with the Gabor scheme of holography is that, because it is the intensity distribution of the hologram which is recorded, the reconstructed image is accompanied by a defocused, aberrated 'conjugate' image which confuses the background. Later, Leith and Upatnieks [59] introduced the idea that, instead of the incident, transmitted wave, a wave passing through vacuum, outside of the object, could act as the reference wave and form a pattern of interference fringes with the wave scattered by the specimen to establish the relative phases of the scattered beams. This opened the way for the major development of holography with light waves as well as other radiations, including high energy electrons, as described below. However, the possible use of the coherent diffraction patterns and shadow-images, or in-line holograms, for image resolution enhancement still has considerable appeal.

In normal CTEM or STEM imaging, one image signal is obtained for each point of the image. However, in a STEM system, a complete two-dimensional array of data, including interference information on relative phases of scattered beams, is given for each incident probe position. In principle, it should be possible to make use of all the information contained in these diffraction patterns to greatly enhance the amount of information derived concerning the specimen. The term ptychology has been introduced to describe various methods by which this might be done [29].

It was pointed out above, that the interference of adjacent overlapping spots in the diffraction pattern of a thin single crystal could give relative phases and so allow structure analysis of the crystal. Konnert et al. [60] proposed a more

sophisticated method by which the diffraction patterns from a series of closely-spaced positions of the incident beam could be Fourier transformed to give a series of correlated real-space autocorrelation functions from which the structure of a crystal and defects in the structure might be derived. This process was carried out to give a reconstruction of the [1 1 0] projection of a Si crystal lattice with a resolution of 0.1 nm or better.

A related approach to the problem has been proposed by Rodenburg and Bates [61]. From Eq. (2), if the diffraction pattern is recorded as a function of the two-dimensional vector, $\boldsymbol{u}$, at all the incident beam positions defined by the two-dimensional vector, $\boldsymbol{X}$, a four-dimensional function is created which, in the projection-function approximation used for Eqs. (4) and (5), may be written

$$\begin{aligned} I(\boldsymbol{u},\boldsymbol{X}) &= |T(\boldsymbol{u})\exp(2\pi i\boldsymbol{u}.\boldsymbol{X}) - P(\boldsymbol{u}) \\ &\quad * T(\boldsymbol{u})\exp(2\pi i\boldsymbol{u}.\boldsymbol{X})|^2 \\ &= |T(\boldsymbol{u})|^2 - \int T^*(\boldsymbol{u})\,P(\boldsymbol{U})\,T(\boldsymbol{u}-\boldsymbol{U}) \\ &\quad \times \exp(-2\pi i\boldsymbol{U}.\boldsymbol{X})\,\mathrm{d}\boldsymbol{U} - \text{c.c.} \quad (10) \end{aligned}$$

Fourier transforming this function with respect to $\boldsymbol{X}$ gives the four-dimensional function,

$$\begin{aligned} G(\boldsymbol{u},\rho) &= |T(\boldsymbol{u})|^2\,\delta(\boldsymbol{r}) - T^*(\boldsymbol{u}) \\ &\quad \times T(\boldsymbol{u}-\rho)\,P(\rho) - \text{c.c.} \quad (11) \end{aligned}$$

Then, because $|T(\boldsymbol{u})|^2 = 1$ if the aperture limitation is neglected, the section of this function represented by $G(\rho/2,\rho)$ is just $\delta(\boldsymbol{u}) - P(\boldsymbol{u}) - P^*(\boldsymbol{u})$, or, in the weak phase-object approximation, the Fourier transform of $1 + 2\sigma\phi(\boldsymbol{X})$.

The validity of this approach has been verified by Rodenburg et al. [62] using light-optical analogue experiments and STEM experiments with moderate resolution. Complications arise, for example, from the fact that the aperture limitation cannot be ignored, but evidence that the method may provide considerable enhancement of resolution has been given.

2.2.7 Holography

The concept of holography can be applied with a STEM instrument in a number of different ways including both in-line and off-axis forms [63]. As pointed out above, the shadow images formed with a stationary incident beam are just the holograms of Gabor's original proposal of the holography concept [56, 57]. If the shadow images are recorded with large defocus and so with moderate magnification, the coherent diffraction effects are visible. In particular, Fresnel fringes occur at any discontinuities and the effects are seen of any deflection of the incident beam by potential field gradients. This is the basis for the Fresnel imaging of the magnetic domain boundaries in thin films of ferromagnetic materials [64]. Reconstruction from such Fresnel images to give the phase variations in the object with high spatial resolution is possible but, as in the original Gabor form of holography, the desired image is confused by an unwanted, defocused conjugate image.

If the beam in a STEM instrument is scanned over the specimen and the diffraction pattern, or near-focus shadow image, is recorded for a closely-spaced set of beam positions, reconstruction, with correction of aberrations, may be made for each recorded pattern and correlation between

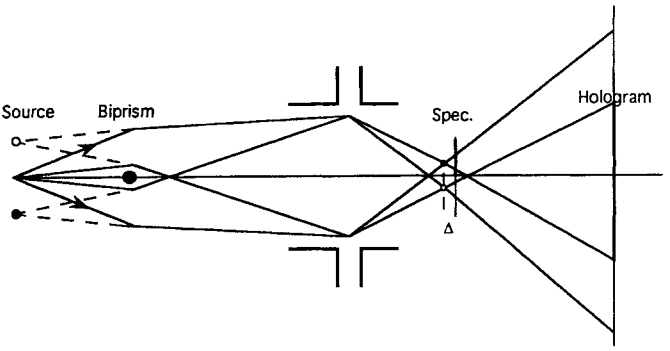


Figure 11. Diagram of the arrangement for off-axis STEM holography. The insertion of an electrostatic biprism near the source produces two coherent effective sources and two probes at the specimen plane, giving interference fringes in the plane of observation.

patterns can be used to reduce, and effectively eliminate, the effects of the conjugate images [55]. An alternative approach is that of Veneklasen [65], who proposed that by making a detector with a detector function $D(u) = T(u)$, an on-line reconstruction of the wave-function may be made, giving a STEM image directly proportional to the projected potential if the weak phase object approximation is valid. Analysis of this scheme suggests that the desired image will be accompanied by a weak, defocused and aberrated conjugate image, which may not be a serious impediment, and that the difficulty of making a detector function having the complex form of $T(u)$ may be avoided [20]. However, this scheme has not yet been realized in practice.

An off-axis form of STEM holography, equivalent in some respects to the off-axis form of CTEM holography (Sec. 1.8 of this Chapter), may be realized, in either stationary-beam or scanning-beam form, if an electrostatic biprism is inserted in the illumination system before the specimen. The biprism has the effect of producing two coherent virtual sources so that two mutually coherent probes are formed at the specimen level, with a variable separation so that one may pass through the specimen and the other may pass through vacuum and act as a reference wave (Fig. 11). The essential difference from the CTEM off-axis holography scheme is that the pattern of interference fringes is formed in the diffraction plane rather than the image plane; but equivalent reconstruction schemes, giving equivalent enhancement of image resolution by correction for aberration effects, are possible [66].

For greatly defocused shadow images, the off-line holography scheme gives a set of interference fringes superimposed on what can be regarded as an image plane and reconstruction of the phase and amplitude distribution of the wave transmitted through the specimen can be made as for the CTEM case. This has been the basis for a very effective means for the imaging and quantitative measurement of the magnetic fields in thin films and around small particles of ferromagnetic materials, as illustrated in Fig. 12 [67]. The spatial resolution of the phase distributions derived from the holograms is limited by the Fresnel diffraction effects if no correction is made for the effects of defocus, but this limited resolution may be sufficient for many observations of domain boundaries. However, if the reconstruction process

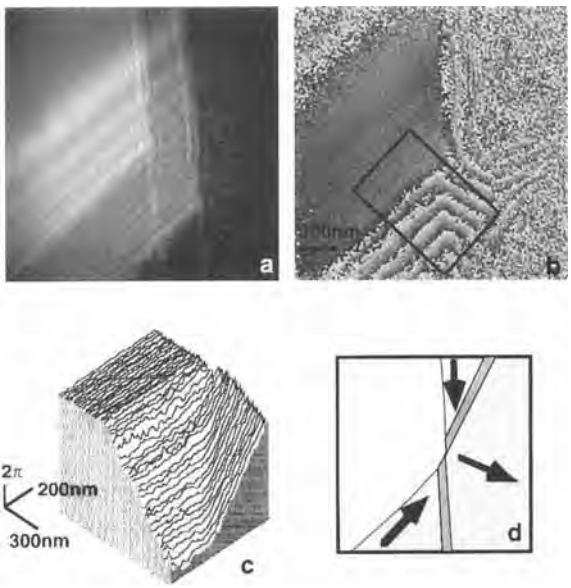


Figure 12. Off-axis hologram in a STEM instrument with a stationary beam obtained from a thin ferromagnetic film (a), and the reconstruction (b), showing the variation of the magnetic field around a domain boundary. The contrast goes from black to white for phase changes of 2π (i.e., it is a 'wrapped' image). (c) An unwrapped image; (d) the magnetic field directions [67].

includes a correction for the defocus effect, the resolution achieved can be 1 nm or better [68].

For the scanning mode in STEM, with a biprism used to produce an interference fringe pattern, a reconstruction of the phase and amplitude components of the specimen wave can be achieved, without the complication of a conjugate image, by use of a special detector configuration such as an elaboration of that proposed by Veneklasen [65].

A much simpler form of detector, consists of a set of fine parallel lines, has been used effectively by Leuthner et al. [69]. The central part of the set of fine interference fringes forming the hologram, or an optically magnified and projected image of it, falls on a grating of black and transparent lines of matched spacing and the transmitted intensity is detected with a photomultiplier. Any variation of the phase of the specimen wavefunction relative to that of the reference wave gives a shift of the fringes and hence a variation of the intensity transmitted through the grating. The averaged intensity of the fringes gives the amplitude of the wave function. This approach has given phase distributions and contoured maps of potential distributions with a spatial resolution equal to that for bright-field imaging of the STEM instrument used.

For all forms of holography, whether in CTEM or STEM, the process of reconstruction to enhance the resolution of the image relies on an accurate knowledge of the imaging parameters. For the current level of attainable image resolution, the most important parameters to be considered are the astigmatism, the defocus and the spherical aberration coefficient. For any improvement of resolution by holographic means, or for the quantitative interpretation of images obtained by any method, the requirements for accurate knowledge of these parameters become very exacting [70]. For STEM, the

observation of shadow images, particularly for the periodic structures of crystals, provides a very convenient means for the correction of astigmatism and the measurement of defocus and spherical aberration.

In underfocused shadow images, as suggested by Fig. 4, there is a well-defined circle of infinite magnification. In particular, for the straight edge of a specimen, this circle is clearly defined and is bisected by the image of the edge with an inversion of the image from the inside to the outside of the circle. The light-optical analogue of this is the well-known 'knife-edge' test used to detect astigmatism of optical lenses. Astigmatism distorts the circle into an S-shaped curve. The astigmatism in the STEM objective lens is corrected by removing such a distortion in the shadow image. An even more stringent test for astigmatism is given by the observation of the Ronchi fringes in the shadow images of thin crystals in that any astigmatism leads to a distortion of the characteristic form of the fringes near focus, illustrated in Fig. 10 [71].

The form of the Ronchi fringes is strongly dependent on both the defocus and C_s. It was pointed out by Lin and Cowley [54] that these parameters can be determined separately by making observations under special conditions. Thus, for the Ronchi fringes from crystal lattice planes of moderately large spacing (~ 0.7–1.0 nm) it is seen that the fringe contrast goes to zero and reverses on a set of concentric ellipses (Fig. 10c). The ratio of the squares of the major axes of any two consecutive ellipses depends only on the spherical aberration constant, C_s, which can therefore be determined independently of the defocus. Once the value of C_s is determined in this way, the defocus can be derived from the shape of the Ronchi fringes and the number of fringes within the infinite magnification circle [55].

2.2.8 STEM Instrumentation

Apart from the early instruments made in individual laboratories, following the original designs of Crewe et al. [1–3], all of the current dedicated STEM instruments are those made by the one commercial manufacturer, VG Microscopes (UK). STEM attachments are available for a number of commercial CTEM instruments but, even when a field-emission gun is provided, the STEM performance tends to be limited by the essential features of the design which are optimized for the CTEM performance. The main imaging field of the objective lens comes after, rather than before, the specimen. Probe sizes at the specimen level of 1 nm or less can be achieved but the intensity within such a probe is usually much less than for the dedicated STEM instrument and insufficient for convenient nanodiffraction or microanalysis. Hence our discussion here will be limited to the one commercial class of dedicated STEM machines.

The cold field emission gun is operated at up to 100 keV (or in a few cases, up to 300 keV) and is placed at the bottom of the column for the sake of mechanical stability. Since the effective source size for a FEG is about 4 nm, it is clear that mechanical vibration of the field-emission tip with an amplitude of 1 nm or less may have an adverse effect on the achievable resolution. Since a large demagnification of the effective electron source is not required, only

two weak condenser lenses are used before the strong objective lens which forms the small probe on the specimen. The condenser lenses allow the choice of a range of probe sizes and intensities to suit the particular application.

The insertion of an objective aperture between the pole pieces of the objective lens, as is done for a CTEM instrument, is often considered undesirable because it limits the space available for specimen manipulators and also may produce unwanted X-ray signals, complicating the interpretation of EDS spectra and X-ray images. Hence, a virtual objective aperture is placed in the space before the condenser lenses.

In many STEM instruments, the electrons passing through the specimen are allowed to diverge until they reach the detector plane, with no lens action except for that of the weak postspecimen field of the asymmetrical objective lens. However, for the convenient observation of diffraction patterns or shadow images with variable magnification, or for holography, it is convenient to have at least two postspecimen lenses. Some recent instruments, in fact, have a symmetrical objective lens so that the postspecimen field of the lens is strong, plus two intermediate lenses and a strong projector lens. The postspecimen lens system is then equivalent to that of a normal CTEM instrument and there is great flexibility for the various detector configurations.

The easy access to the top of the STEM column provides the necessary flexibility in the construction of the detector system. The minimum detection system includes a phosphor or scintillator screen for observation of the diffraction pattern, recorded by means of a low light-level TV camera with a VCR, and preferably a similar, alternate screen with an aperture which allows a selected part of the diffraction pattern to pass through to the EELS spectrometer for bright-field or dark-field imaging with filtered, elastically scattered electrons, for imaging with selected energy-loss electrons, or for EELS microanalysis of selected regions of the specimen. In addition there should be one, or preferably several, interchangeable, annular detectors to provide flexibility in ADF, HAADF imaging or special bright-field imaging modes.

For quantitative recording of diffraction patterns, shadow-images or holograms, the scintillator-TV combination is replaced, interchangeably by a CCD camera system. With current CCD systems the scan rate is slow so that one image frame is recorded in one second or often longer and no provision is made for switching to a fast scan. Since fast scans at TV rates are essential while searching for the desired specimen area, or for focusing and stigmating, it is important to be able to switch quickly from a TV to the CDD detector. Provision is also desirable for introducing specially shaped detectors or masks to implement the several suggested imaging modes involving detectors of non-standard configuration, such as those of Leuthner et al. [69] or Veneklasen [65]. A detector system designed to have all these desirable features with reasonably efficient signal collection has recently been described [20] but is undergoing redesign for further improvement.

The usual asymmetrical design for the STEM objective lens has some advantages in allowing easier access to the specimen region for the EDS X-ray detector, which, for efficiency of signal collection, needs to

be as close to the specimen as possible. Also it is relatively easy with this lens configuration to hold specimens in the space outside the strong magnetic field of the objective lens so that their magnetic configurations can be observed under controlled conditions of externally applied magnetic field [64, 67].

The signal-handling capabilities needed for STEM have been well-developed in many cases. Digital scans may be substituted for the conventional analog scans. The images produced in the scanning mode may be displayed directly on cathode ray tubes and recorded with parallel photographic monitors. Signals from several detectors may be displayed and recorded simultaneously, or may be combined by addition, subtraction or multiplication and recorded for special purposes. Alternatively, the signals may be digitized and recorded for further processing, on-line or off-line, by attached computer systems. To overcome the problem of a fluctuating emission from the electron gun to some extent, a reference signal may be obtained from, for example, the electrons striking the objective aperture blade but not transmitted: then the ratio of the image signal to this reference signal is recorded.

For the stable operation of a cold field-emission gun, the vacuum in the gun chamber must be better than 10^{-10} torr and the column vacuum must be 10^{-8} or better to prevent excessive backstreaming of gas molecules into the gun. The STEM instruments have a column vacuum of better than 10^{-8} for this reason and also in order to minimize the effects of contamination of the specimen. For poorer vacuum, or with specimens which have not been adequately cleaned, the contamination rate can be very high when the incident beam illuminates only a very small part of the specimen. Then the migration of organic molecules along the specimen surface is not prevented by the flooding of a large specimen area around the area of interest by a broad electron beam, as is common in CTEM. With care, however, in a STEM instrument the contamination rate can be reduced to a level for which no accumulation of contaminant is observable, even with a focused probe, over a period of many hours.

With special care in the construction and operation, STEM instruments can have a column vacuum of better than 10^{-9} torr, and in at least one case, an instrument has been made with a completely UHV system, with better than 10^{-10} in the whole system, including an elaborate auxiliary chamber for specimen preparation and manipulation [72]. This instrument, the so-called MIDAS system (a microscope for imaging, diffraction and analysis of surfaces), was specifically designed for research on surface structures with high spatial resolution. The various STEM imaging modes, nanodiffraction and EELS microanalysis can be performed on surfaces and thin films which are clean to the exacting standards demanded in surface science.

In addition, SEM and SAM signals can be collected and analyzed with respect to their energy distributions, with high spatial and energy resolutions, through the application of magnetic 'parallelizers'. The low-energy electrons emitted from the sample in the high magnetic field of the objective lens spiral around the lens axis as they drift out of the field. If the field is properly shaped, the spiral angle of the electrons decreases with the magnetic field (i.e., the electrons are parallelized) and

when they emerge into field-free space they are confined to a cone a few degrees wide, ideally suited for feeding into a low-energy electron spectrometer for energy analysis and detection with high efficiency [73]. The spatial resolution for SAM images obtained in this way has reached the level of 1 nm, an order of magnitude better than for other instruments [44].

As in the case of CTEM, images of the surfaces of crystals can be obtained in STEM instruments if a bulk specimen is mounted so that the incident beam strikes an extended, nearly flat surface at a grazing angle of incidence. The diffraction pattern formed on the detector plane is then a convergent beam reflection high energy electron diffraction (RHEED) pattern, similar to a normal RHEED pattern except that the individual diffraction spots are replaced by diffraction disks. If a region of high intensity within one of the strong RHEED spots is selected to give the STEM signal, the scanning reflection electron microscopy (SREM) image produced resembles the corresponding REM image formed in a CTEM instrument (see Sec. 1.2 of this Chapter).

It is interesting to note that the highest intensity regions within an extended convergent-beam reflection high-energy electron diffraction (CBRHEED) spot often correspond to incident beam orientations for which strong surface resonance effects occur. As in REM, the image intensity is highly sensitive to small defects on the crystal surface, such as steps one atom high or more, the strain fields of emerging dislocations or other crystal faults, and variations of the structure or composition of the topmost one or two layers of atoms on the surface [74]. As compared with REM, SREM has the advantage that SEM and SAM signals are produced at the same time, or subsequently with higher angles of incidence, to give complementary information on the surface morphology and composition. The only modification required for a STEM instrument to be used for SREM is that the specimen should be mounted with an extended flat face nearly parallel to the incident beam.

2.2.9 Applications of Scanning Transmission Electron Microscopy

Although various applications of STEM will be included in Volume 2 of this Handbook, a few will be mentioned here to illustrate the capabilities of the technique and also illustrate the nature of the information that it can provide.

In the initial work by Crewe's group [75] it was clearly established that it was possible with ADF imaging to obtain clear images of individual heavy atoms on light atom supports (Fig. 1). The heavy atoms could be as light as Ag [76]. Several possible applications immediately suggested themselves. It was considered, for example, that if heavy atoms were attached to particular known sites on biological macromolecules it would be possible to deduce something about the structures of the molecules from the images of the heavy atoms even if the molecules themselves were destroyed by radiation damage in the course of the imaging. Some limited success was achieved in this endeavor, particularly by Beers et al. [77]. However, questions arose as to whether the heavy atoms could remain in their original sites

while the molecules around them disintegrated.

In observations of heavy atoms sitting on thin films of amorphous carbon, it appeared that the heavy atom positions tended to change from one image scan to the next [78]. Experiments were made to determine whether the atom movements resulted from energy pulses given to the atoms by the incident electron beam by inelastic scattering events, or whether they derived from random thermal vibration excitations. The conclusion was that the latter was more likely.

The quantitative nature of the ADF image signal in a STEM instrument gave rise to the suggestion that measurements of image intensities could be used to determine the masses of small particles and, in particular, of individual macromolecules [79]. To a good approximation, the intensity collected by the annular detector is proportional to the total scattering from a nearly amorphous object, since the intensity in the diffraction pattern falls off smoothly with scattering angle. The mass per unit area of the specimen could be derived from measurements of the scattered intensity by comparison with the scattering cross section per unit mass, derived from the theoretical atomic scattering cross sections. Measurements made on particles of known structure gave good agreement with the masses derived from other methods. On this basis, a large number of valuable mass determinations for a wide variety of biological particulates have been made, particularly by Wall (who initially proposed the method) and associates [80]. See also [81].

A widely explored application of STEM has been for the study of small particles and particularly of the small particles of heavy metals in light-atom supports, such as the active components in supported metal catalysts. These particles have sizes typically in the range 1–10 nm. When such particles are embedded in the near-amorphous support material, or even when they are placed on the usual amorphous carbon supporting films, it is difficult to measure, or even detect them in bright-field CTEM or STEM images because the phase-contrast noise from the amorphous material gives random intensity fluctuations on much the same scale. The ADF, and especially the HAADF, imaging mode can reveal such particles clearly even when they are completely obscured in BF images by relatively large thickness of the support material.

Once a particle is detected in an HAADF STEM image, it may be selected for analysis by nanodiffraction, EELS or EDS, when the incident beam is stopped at that point. In addition, in specially equipped instruments, the HAADF image may be correlated with SEM or SAM images of the same specimen area to determine whether the particle is on the top or bottom surface of the sample or in its interior. This information is often of importance in that it can indicate the extent to which the catalyst particle may be exposed to the reacting gaseous atmosphere. Figure 13 shows SEM images from the top side and bottom side of a sample of gold particles on a thin carbon support [43]. Transmission images give no indication as to the relative positions of the gold particles and the carbon film in the beam direction. The SEM images, obtained simultaneously with detectors above and below the specimen, indicate clearly that the gold particles are all on the bottom side of the supporting film, as are some

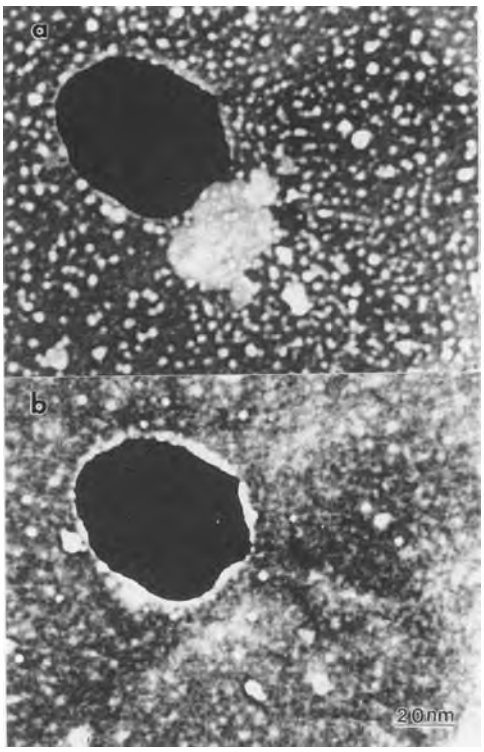


Figure 13. SEM images obtained from the entrance (a) and exit (b) surfaces of a carbon film with small metal particles and some light-atom material on the entrance side (Courtesy of J. Liu [96]).

light-atom particles, visible in one image but not the other. It may be noted that weak, diffuse images of the gold particles appear in the SEM images even on the 'wrong' side of the thin carbon film, presumably because some of the low-energy secondary electrons can penetrate through the film.

The problem of locating heavy-metal atoms or small clusters within the channels in the structures of zeolites becomes more significant as the use of zeolites in industrial catalysts becomes more extensive. A major difficulty for the application of electron microscopy techniques to such materials is that the zeolite structural framework is rapidly destroyed by electron irradiation. It is necessary to make use of minimum exposure techniques so that the image information is obtained from the first electrons to strike the sample area of interest. This can readily be achieved in STEM for which only the area of the specimen actually being scanned is exposed to the electron beam. Initial location of interesting specimen areas can be done at low magnification with low irradiation doses. Final focusing and stigmation can be done on adjacent areas before the beam is switched to the region of interest where a single scan is made and used to record the image. A comparison of CTEM and HAADF STEM imaging by Rice et al. [82] showed the latter to be more effective in revealing few-atom clusters or small metal particles in zeolites. An additional advantage is that nanodiffraction patterns from the particles could be made immediately after the image, to give some information on the relative orientations of the small metal crystals and the walls of the zeolite channels [83].

In Fig. 14, an EDS image obtained with the characteristic X-rays from small Pd particles on a carbon support [84], it is evident that particles as small as 2–5 nm in diameter may be detected, especially when a digital imaging technique is applied. With the intrinsically better detection efficiencies of electron energy-loss imaging, even smaller particles may be imaged [85]. In addition, for very small regions, it is also possible to detect and make use of the fine structure of the energy-loss peaks which is characteristic of the particular valence states or bonding configurations of the atoms [86] (Sec. 1.3 of this Chapter).

Small particles of metals are often single crystals. However, there are many CTEM

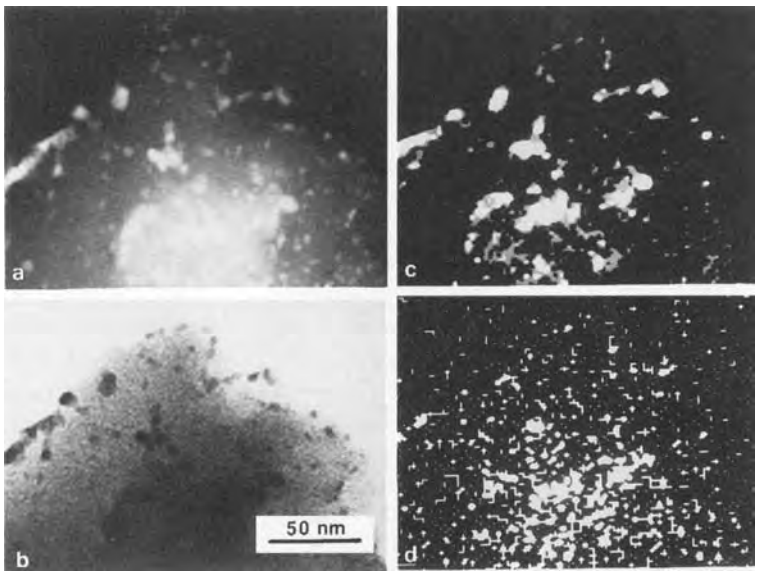


Figure 14. (a) Annular dark-field dedicated STEM image of Pd particles on a carbon support; (b) bright-field STEM image; (c) Pd L-α X-ray image with background X-ray intensity subtracted; (d) background X-ray image of the same area with signal intensity expanded ten times more than (c). Digital images (c) and (d) were smoothed, which caused line artifacts in the low-intensity image (d). Original magnification = 400 000×. (Courtesy of C. E. Lyman [84]).

observations to suggest that particles of noble metals, and particularly of gold, in the size range 10–100 nm are often twinned or multiply twinned (see, for example, Allpress and Sanders [87]). They may form decahedral particles, with five tetrahedrally shaped regions of perfect crystal related by twinning on (1 1 1) planes, or icosahedral particles with twenty tetrahedral, mutually twinned regions. Considerations of the energetics of the small particles suggest that the multiply twinned forms may be the equilibrium forms for very small particles, but the evidence from CTEM is inconclusive for particle sizes below about 10 nm. For one sample, formed by co-sputtering of gold and polyester, the STEM images showed a range of particle sizes down to about 2 nm or less. Nanodiffraction from the individual particles revealed that for 3–5 nm particles about half were multiply-twinned, but the proportion was even less for smaller particles in the range 1.5–3 nm [88]. This result was not necessarily in disagreement with the theoretical studies which referred only to isolated particles with no perturbation of surface energies by support material.

The possibility of recording nano-diffraction patterns at TV rates by using a low light-level TV camera and VCR has allowed detailed studies of several small-scale structures. The structure of the individual planar defects in diamond, thought to be associated with the aggregation of nitrogen atoms, was investigated in this way by Cowley et al. [89]. Patterns were recorded at intervals of 0.02 nm as a beam of 0.3 nm diameter was scanned across a defect with an instrumental magnification of 2×10^7. The observed intensities were compared with dynamical diffraction calculations based on the various models which had been proposed for the defect. The best fit was obtained for a model due to Humble [90] which contained no nitrogen atoms.

Similar series of nanodiffraction patterns were obtained with a beam of

diameter about 1 nm scanned across various carbon nanotubes, formed in a carbon arc discharge in high-pressure He and shown by Iijima [91] to be made up of concentric cylinders of graphitic sheets with internal diameters from about 2 to 10 nm and external diameters from 3 to 20 nm (Fig. 7). It was known that the individual sheets had a helical structure and several different helical angles could be included in one nanotube. It had been assumed that the cylinders had circular cross section. Because a regular graphite crystal structure cannot be consistent with a superposition of graphite sheet cylinders with regularly increasing circumference, it was concluded that the lateral arrangement of the sheets must be disordered. From the nanodiffraction patterns it was evident that, for many nanotubes, the cylinders were not circular but polygonal, and probably pentagonal, in cross section [92]. It was seen that, in the regions of the flat faces of the cylinders, the graphitic sheets were ordered as in the crystalline graphite structure. The nanotubes having polygonal cross section included non-helical ones and ones having one or more helical angles.

The scanning reflection electron microscopy (SREM) mode has been applied in various studies of surface structure (Liu and Cowley [74]). Single-atom-high surface steps and the strain fields of emerging dislocations have been observed with the same characteristic contrast variations with diffraction conditions as in REM (see Sec. 1.2 of this Chapter). Long-period surface superstructures on oxygen-annealed cleavage faces of MgO have been observed and studied (Fig. 15). Interesting differences in image contrast have been seen for SREM and the allied

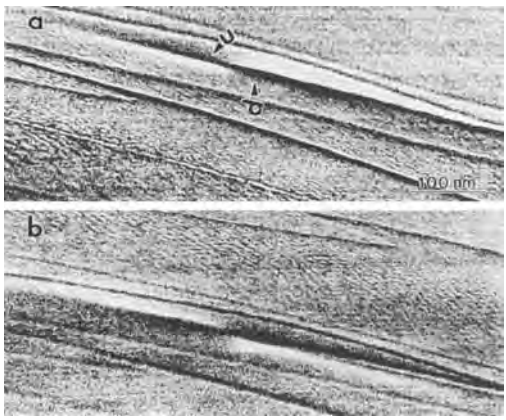


Figure 15. Scanning reflection electron microscopy (SREM) image of the (1 0 0) face of MgO crystal showing the fringes due to a surface superlattice in two directions and the reversal of contrast of the streaks due to the strain field around an emerging dislocation as the defocus is reversed from (a) to (b). (Courtesy of J. Liu [74]).

techniques of SEM imaging and the SREM imaging mode using a high-angle annular detector, in analogy with the transmission HAADF mode [74]. For the latter two modes there is little dependence of the image contrast on diffraction conditions and the image shows mostly the surface morphology with different characteristic intensity variations in the two cases.

As in the case of STEM, the SREM technique has the virtue that the beam scan may be stopped at any point of the image so that nanodiffraction or microanalysis with EELS or EDS may be performed on any chosen feature. In studies of the deposition of Pd metal on MgO surfaces, for example, nanodiffraction patterns from individual particles of about 2 nm diameter sitting on the surface showed them to be single crystals of Pd but, under continued electron beam irradiation, these crystals were seen to be gradually covered by a layer of different

material, identified by nanodiffraction as single crystal PdO [93]. The SREM technique has also been used by Milne [94] for studies of the oxidation of copper surfaces. The images revealed the form of the copper oxide crystallites growing on the copper surfaces and their nucleation sites relative to the steps and facets of the copper crystal cut at a small angle to the (1 0 0) surface and annealed to give a surface reconstruction.

It is evident that, for both transmission and reflection modes, the STEM instruments have particular capabilities which make them invaluable for some special purposes, particularly when the combination of imaging, nanodiffraction and microanalysis provides a much more complete assessment of a specimen than can be obtained by any other approach. The full exploitation of the coherent interference effects observable in the diffraction patterns remains as a challenge for the future.

2.2.10 References

[1] A. V. Crewe, J. Wall, *J. Mol. Biol.* **1970,** *48*, 373.
[2] A. V. Crewe, *Rep. Progr. Phys.* **1980,** *43*, 621.
[3] A. V. Crewe, J. Wall, J. Langmore, *Science* **1970,** *168*, 1333.
[4] J. M. Cowley, *Appl. Phys. Letts.* **1969,** *15*, 58.
[5] E. Zeitler, M. G. R. Thompson, *Optik* **1970,** *31*, 258, 359.
[6] J. R. Sellar, J. M. Cowley in *Scanning Electron Microscopy 1973* (Ed.: O. Johari), IIT Research Institute, Chicago **1973,** p. 243.
[7] T. Groves, *Ultramicroscopy* **1975,** *1*, 15.
[8] A Strojnik in *Microscopie Electronique à Haute Tension* (Eds: B. Jouffrey, P. Favard), Société Française de Micros. Electronique, Paris **1976.**
[9] J. M. Cowley, *Ultramicroscopy* **1976,** *2*, 3.
[10] A. Howie, *J. Microsc.* **1979,** *17*, 11.
[11] M. M. J. Treacy, A. Howie, *J. Catal.* **1980,** *63*, 265.
[12] Z. L. Wang, J. M. Cowley, *Ultramicroscopy* **1990,** *32*, 275.
[13] P. Zu, R. F. Loane, J. Silcox, *Ultramicroscopy* **1991,** *38*, 127.
[14] S. J. Pennycook, D. E. Jesson, *Ultramicroscopy* **1991,** *37*, 14.
[15] S. Hillyard, J. Silcox, *Ultramicroscopy* **1993,** *52*, 325.
[16] N. H. Dekkers, H. de Lang, *Optik* **1974,** *41*, 452.
[17] J. M. Chapman, P. E. Batson, E. M. Waddell, R. P. Ferrier, *Ultramicroscopy* **1978,** *3*, 203.
[18] I. R. McFayden, J. M. Chapman, *Electron Microsc. Soc. Amer. Bull.* **1992,** *22*, 64.
[19] M. Hammel, H. Kohl, H. Rose in *Proc. XIIth Internat. Congress Electron Microsc.* (Eds.: L. D. Peachey, D. B. Williams) San Francisco Press, San Francisco **1990,** Vol. 1, p. 120.
[20] J. M. Cowley, *Ultramicroscopy* **1993,** *49*, 4.
[21] J. M. Cowley, A. Y. Au in *Scanning Electron Microscopy, Vol. 1* (Ed.: O. Johari), SEM Inc., Chicago **1978,** p. 53.
[22] J. Liu, J. M. Cowley, *Ultramicroscopy* **1993,** *52*, 335.
[23] H. Rose, *Ultramicroscopy* **1977,** *2*, 251.
[24] W. Kunath, H. Gross, *Ultramicroscopy* **1985,** *16*, 349.
[25] C. J. Humphreys, E. G. Bithell in *Electron Diffraction Techniques, Vol. 1* (Ed.: J. M. Cowley), Oxford University Press **1992,** p. 75.
[26] J. Barry in *Electron Diffraction Techniques, Vol. 1* (Ed.: J. M. Cowley), Oxford University Press **1992,** p. 170.
[27] J. M. Cowley, J. C. H. Spence, *Ultramicroscopy* **1981,** *6*, 359.
[28] J. C. H. Spence, J. M. Cowley *Optik* **1978,** *50*, 129.
[29] W. Hoppe, *Ultramicroscopy* **1982,** *10*, 187.
[30] R. Nathan in *Digital Processing in Biomedical Imaging* (Eds.: K. Preston, M. Onoe), Plenum Press, New York **1976,** p. 75.
[31] P. E. Batson, *Ultramicroscopy* **1992,** *47*, 133.
[32] F. P. Ottensmeyer, J. W. Andrew, *J. Ultrastruct. Res.* **1980,** *72*, 336.
[33] D. B. Williams, J. R. Michael, J. I. Goldstein, A. D. Romig, Jr. *Ultramicroscopy* **1992,** *47*, 121.
[34] M. Scheinfein, A. Muray, M. Isaacson, *Ultramicroscopy* **1985,** *16*, 237.
[35] H. Raether, *Excitation of Plasmons and Interband Transitions by Electrons, Springer Tracts in Modern Physics*, Vol. 88, Springer, New York **1980.**
[36] Z. L. Wang, J. M. Cowley, *Ultramicroscopy* **1987,** *21*, 77.
[37] J. M. Cowley, *Phys. Rev. B* **1982,** *25*, 1401.
[38] N. Zabala, P. M. Echenique, *Ultramicroscopy* **1990,** *32*, 327.

[39] R. H. Milne, G. G. Hembree, J. S. Drucker, C. J. Harland, J. A. Venables, *J. Microsc.* **1993,** *170*, 193.
[40] K. Kuroda, S. Hosoki, T. Komoda, *Scanning Microscopy* **1987,** *1*, 911.
[41] J. Liu, J. M. Cowley, *Scanning Microscopy* **1988,** *2*, 1957.
[42] M. R. Scheinfein, J. S. Drucker, J. K. Weiss, *Phys. Rev. B* **1993,** *47*, 4068.
[43] G. G. Hembree, P. A. Crozier, J. S. Drucker, M. Krishnamurthy, J. A. Venables, J. M. Cowley, *Ultramicroscopy* **1989,** *31*, 111.
[44] J. Liu, G. G. Hembree, G. E. Spinnler, J. A. Venables, *Ultramicroscopy* **1993,** *52*, 369.
[45] J. M. Cowley, *Ultramicroscopy* **1979,** *4*, 435.
[46] W. J. Vine, R. Vincent, P. Spellward, J. W. Steeds, *Ultramicroscopy* **1992,** *41*, 423.
[47] J. M. Cowley in *Scanning Electron Microscopy 1980, Vol. 1* (Ed.: Om Johari), SEM Inc., Chicago **1980,** p. 61.
[48] J. Zhu, J. M. Cowley, *Acta Crystallogr. A* **1982,** *38*, 718.
[49] J. Zhu, J. M. Cowley, *J. Appl. Crystallogr.* **1983,** *16*, 171.
[50] M. Pan, J. M. Cowley, J. C. Barry, *Ultramicroscopy* **1989,** *30*, 385.
[51] J. M. Cowley, *Sov. Phys. Crystallogr.* **1981,** *26*, 549.
[52] J. M. Cowley, *Ultramicroscopy* **1979,** *4*, 413.
[53] V. Ronchi, *Appl. Optics* **1964,** *3*, 437.
[54] J. A. Lin, J. M. Cowley, *Ultramicroscopy* **1986,** *19*, 31.
[55] S.-Y. Wang, J. M. Cowley, *Micros. Res. Techs.* **1995,** *30*, 181.
[56] D. Gabor, *Nature* **1948,** *161*, 777.
[57] D. Gabor, *Proc. R. Soc. Lond.* **1949,** *A197*, 454.
[58] J. A. Lin, J. M. Cowley, *Ultramicroscopy* **1986,** *19*, 179.
[59] E. N. Leith, J. Upatnieks, *J. Opt. Soc. Amer.* **1962,** *52*, 1123.
[60] J. Konnert, P. D'Antonio, J. M. Cowley, A. Higgs, H. J. Ou, *Ultramicroscopy* **1989,** *30*, 371.
[61] J. M. Rodenburg, R. H. T. Bates, *Philos. Trans. R. Soc. Lond.* **1992,** *A339*, 521.
[62] J. M. Rodenburg, B. C. McCallum, P. D. Nellist, *Ultramicroscopy* **1993,** *48*, 304.
[63] J. M. Cowley, *Ultramicroscopy* **1992,** *41*, 335.
[64] I. R. McFayden, J. N. Chapman, *Electron Micros. Soc. Amer. Bull.* **1992,** *22*, 64.
[65] L. H. Veneklasen, *Optik* **1975,** *44*, 447.
[66] M. A. Gribelyuk, J. M. Cowley, *Ultramicroscopy* **1992,** *45*, 115.
[67] M. Mankos, M. R. Scheinfein, J. M. Cowley, *J. Appl. Phys.* **1994,** *75*, 7418.
[68] M. Mankos, P. de Haan, V. Kambersky, G. Matteucci, M. R. McCartney, Z. Yank, M. Scheinfein, J. M. Cowley in *Electron Holography* (Eds.: A. Tonomura et al.), Elsevier Science BV **1995,** p. 329.
[69] Th. Leuthner, H. Lichte, K.-H. Herrmann, *Phys. Stat. Solidi (a)*, **1989,** *116*, 113.
[70] H. Lichte, *Ultramicroscopy* **1993,** *51*, 15.
[71] J. M. Cowley, M. Disko, *Ultramicroscopy* **1980,** *5*, 469.
[72] J. A. Venables, J. M. Cowley, H. S. von Harrah in *Inst. Phys. Conf. Series, No. 90, Chapter 4*, IOP Publishing, Bristol **1987,** p. 81.
[73] P. Kruit, J. A. Venables, *Ultramicroscopy* **1988,** *25*, 183.
[74] J. Liu, J. M. Cowley, *Ultramicroscopy* **1993,** *48*, 381.
[75] J. P. Langmore, J. Wall, M. S. Isaacson, *Optik* **1973,** *38*, 335.
[76] M. Retsky, *Optik* **1974,** *41*, 127.
[77] M. D. Cole, J. W. Wiggins, M. Beer, *J. Mol. Biol.* **1978,** *117*, 378.
[78] M. Isaacson, D. Kopf, M. Utlaut, N. W. Parker, A. V. Crewe, *Proc. Nat. Acad. Sci. USA* **1977,** *74*, 1802.
[79] M. K. Lamvik, J. P. Langmore, *Scanning Electron Microscopy 1977, Vol. 1* (Ed.: Om Johari), IIT Research Institute, Chicago **1977,** p. 401.
[80] A. V. Crewe, J. Wall in *Proc. 29th Annual Meeting, Electron Micros. Soc. Amer.* (Ed.: C. J. Arceneaux), Claitors Publ. Divis., Baton Rouge **1971,** p. 24.
[81] G. J. Brakenhoff, N. Nanninga, J. Pieters, *J. Ultrastructure Research* **1972,** *41*, 238.
[82] S. B. Rice, J. Y. Koo, M. M. Disko, M. M. J. Treacy, *Ultramicroscopy* **1990,** *34*, 108.
[83] M. Pan, J. M. Cowley, I. Y. Chan, *Ultramicroscopy* **1990,** *34*, 93.
[84] C. E. Lyman, *Ultramicroscopy* **1986,** *20*, 119.
[85] E. Freund, J. Lynch, R. Szymansky, *Ultramicroscopy* **1986,** *20*, 107.
[86] P. Rez in *Physical Methods of Chemistry, Vol. 4, Microscopy* (Eds.: B. W. Rossiter, J. F. Hamilton), John Wiley & Sons, New York **1991,** p. 203.
[87] J. A. Allpress, J. V. Sanders, *Surf. Sci* **1965,** *7*, 1.
[88] W. B. Monosmith, J. M. Cowley, *Ultramicroscopy* **1984,** *12*, 177.
[89] J. M. Cowley, M. A. Osman, P. Humble, *Ultramicroscopy* **1984,** *15*, 311.
[90] P. Humble, *Proc. R. Soc. Lond.* **1982,** *A381*, 65.
[91] S. Iijima, *Nature* **1991,** *354*, 56.
[92] M. Liu, J. M. Cowley, *Ultramicroscopy* **1994,** *53*, 333.
[93] H. J. Ou, J. M. Cowley, *Phys. Status Solidi (a)* **1988,** *107*, 719.

[94] R. H. Milne in *Reflection High Energy Electron Diffraction and Reflection Electron Imaging of Surfaces* (Eds.: P. K. Larsen, P. J. Dobson), Plenum Press, New York and London **1988,** p. 317.
[95] J. M. Cowley, *Ultramicroscopy* **1989,** *27*, 319.
[96] J. Liu, J. M. Cowley, *Ultramicroscopy* **1991,** *37*, 50.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

2.3 Scanning Transmission Electron Microscopy: *Z* Contrast

2.3.1 Introduction

Z-contrast scanning transmission electron microscopy (STEM) realizes the alluring goal of direct structure retrieval through its incoherent, compositionally sensitive image, at a present resolution of 1.26 Å. Furthermore, it facilitates atomic resolution microanalysis, as the STEM probe can be positioned over known atomic columns for the collection of X-ray or electron energy loss spectroscopy (EELS) data. This powerful combination of column-by-column imaging and spectroscopy opens up many important but previously inaccessible materials problems for fundamental study. Such investigations require incoherent imaging conditions to be established for both the imaging and the analytical signals. In this section, we examine the theoretical basis for incoherent imaging with the three distinct classes of electrons, coherent scattering, thermal diffuse scattering, and inelastically scattered electrons. The power of this approach to structure determination is illustrated through a variety of applications; determining interface and grain boundary structures in semiconductors and ceramics, imaging the morphology of nanocrystalline materials, and elucidating the atomic scale structure of highly dispersed supported catalysts.

An incoherent image is defined as the convolution of two positive quantities, one representing the specimen, which we refer to as the object function $O(\boldsymbol{R})$, and one representing the resolution of the imaging system, which in our case is the probe intensity profile $P^2(\boldsymbol{R})$. The transfer function is then the Fourier transform of the probe intensity profile, which is also generally positive. In Fig. 1, the optimum contrast transfer functions for coherent and incoherent imaging conditions are compared, assuming a 300 kV microscope with an objective lens C_s of 1 mm. The difference between the transfer function of the incoherent mode and the oscillating transfer function of the coherent mode is evident. Because $P^2(\boldsymbol{R})$ is always sharper than $P(\boldsymbol{R})$, the second important property of incoherent imaging is its substantially enhanced resolution; the Scherzer resolution condition for incoherent imaging is $0.43C_s^{1/4}\lambda^{3/4}$, significantly higher than that for coherent imaging, $0.66C_s^{1/4}\lambda^{3/4}$ [1].

These resolution criteria are deduced assuming the objective aperture can be set to the optimum size appropriate to the imaging conditions used. For a

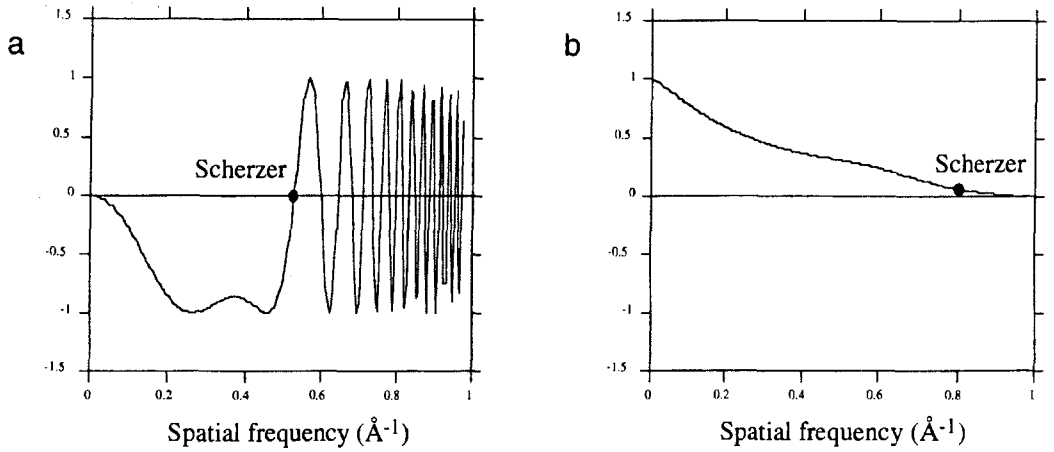


Figure 1. Contrast transfer functions for a 300 kV microscope with a 1 mm C_s objective lens for (a) coherent and (b) incoherent imaging conditions. Curves assume Scherzer optimum conditions: (a) defocus −505 Å, (b) defocus −438 Å, aperture cut-off 0.935 Å^{-1}.

specific aperture size, incoherent dark field imaging gives *double* the resolution of bright field coherent imaging, as shown in Fig. 2. Image contrast results from the regions of overlapping disks that reach the respective detector. In a bright field image, spacings must be below the objective aperture radius for the interference region to fall on the small axial detector, exactly as in conventional transmission electron microscopy (TEM). With an annular detector, interference is detected from

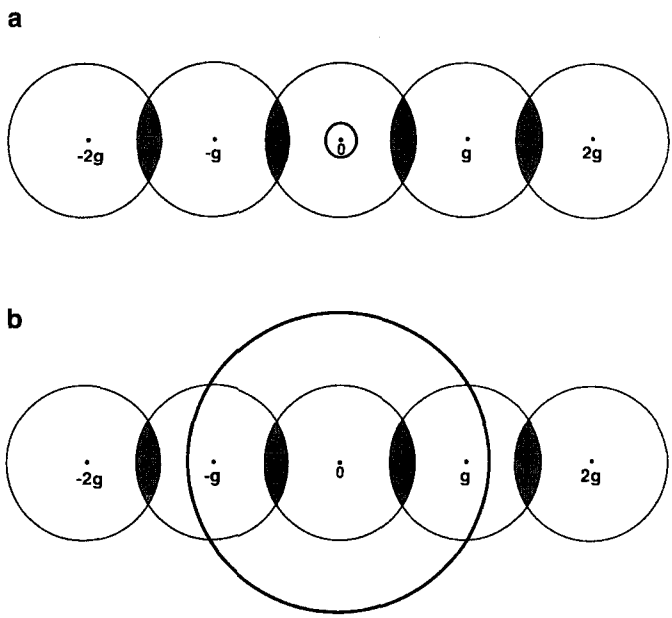


Figure 2. Regions of overlapping convergent beam disks for a diffraction vector greater than the objective aperture radius. (a) An axial bright field detector shows no contrast, while (b) regions of overlapping disks fall on the annular detector. The incoherent dark field image has double the resolution of the bright field image.

spacings up to the aperture diameter. The dark field image therefore shows double the resolution of the bright field image.

A less obvious but equally important advantage of incoherent imaging is that its intuitive relation to the object is maintained under dynamic diffraction conditions. Modern computer-based approaches to phase contrast imaging are successfully extending resolution beyond the Scherzer condition, but direct inversion will be limited to thin objects in which dynamical effects are small. In general, coherent imaging methods must rely on the use of preconceived structure models and 'goodness of fit' measures. Incoherent images are highly intuitive; an atomic column does not reverse contrast even if its image overlaps that of a neighboring column. Column intensities do not increase linearly with column length, due to absorption effects or some partial coherence, but column *locations* can still be determined to high accuracy by direct inspection of the image.

The difference between the two imaging modes is perhaps more apparent in Fig. 3, comparing coherent and incoherent imaging optics in a conventional TEM geometry. Phase contrast imaging (Fig. 3a) involves the coherent transfer of a plane wave fast electron through the specimen and objective lens system to the detector, at which point the intensity is taken. To determine the specimen potential requires first the reconstruction of the exit face wave function, and, second, a solution of the inverse dynamic diffraction problem. Usually, therefore, interpretation must be made through simulations of model structures, with the unavoidable risk that the correct model may not be considered.

These problems are overcome if the coherence of the imaging process is broken, as in Fig. 3b. This requires illumination over a large range of scattering angles, which, through the increased contribution of nuclear Rutherford scattering, is also the reason for the strong atomic number (Z) sensitivity of the images. Then the specimen is made effectively self-luminous, and the same imaging optics now provide an incoherent image with

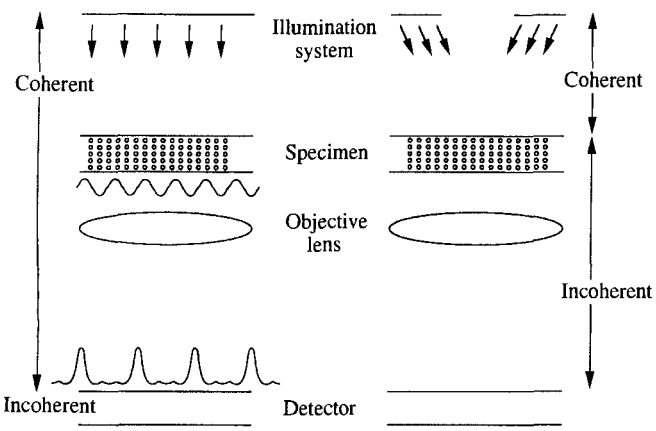


Figure 3. Schematic comparing (a) coherent and (b) incoherent imaging optics for a transmission electron microscope.

the transfer characteristics of Fig. 1b. Contrast reversals can theoretically occur in the presence of aberrations [2], but are rare, and occur only at low amplitude. Therefore, strong image features that do not reverse contrast on changing objective lens defocus can be directly interpreted as the positions of atomic columns. There are no proximity effects at interfaces, no contrast oscillations with increasing thickness, and beam broadening is reduced. There is no phase problem associated with incoherent imaging, and no problem with dynamical diffraction. The reliance on preconceived structure models is removed and direct structure inversion becomes a reality; quantitative methods such as maximum entropy [3] allow atomic column coordinates to be retrieved with an accuracy of a few tenths of an angstrom.

The need for a wide range of scattering angles means that incoherent imaging is more efficiently achieved with the reciprocally equivalent STEM geometry shown in Fig. 4, where the annular detector collects elastically and thermally scattered electrons. Most inelastically scattered electrons pass through the central hole, and can therefore be detected *simultaneously*, which was one of the important motivations for the original development of the scanning transmission electron microscope. Indeed, the first Z-contrast images were obtained by taking the ratio of elastic to inelastic signals [4]. Such methods proved useful in biological studies, but were unsuitable in materials science due to the need for zone axis crystals to achieve atomic resolution, and the unavoidable presence of strong dynamical scattering in most samples of practical interest. How this is overcome represents a fascinating study in the effects of quantum mechanical superposition, as discussed in Sec. 2.3.5 of this Chapter. It is the preferential selection of tightly bound s-type Bloch states that leads to the strong columnar channeling and makes incoherent imaging possible in thick crystals. Thin crystal, weak scattering results can be simply extended into the thick crystal regime. First, therefore, we examine the theory of incoherent imaging in thin crystals using electrons scattered by coherent, thermal, and inelastic mechanisms.

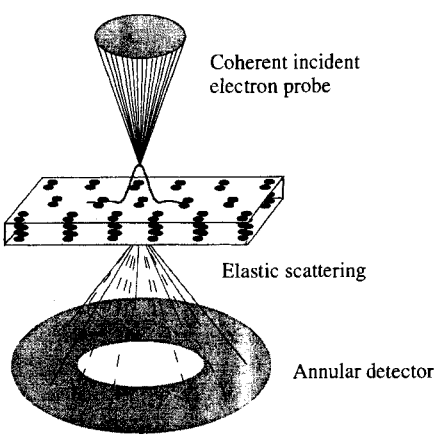


Figure 4. Schematic of the STEM optics for incoherent imaging.

2.3.2 Incoherent Imaging with Elastically Scattered Electrons

For clarity, we will retain the general form $P(\boldsymbol{R} - \boldsymbol{R}_0)$ of the STEM probe amplitude profile, where $\boldsymbol{R}$ is a transverse positional coordinate on the specimen entrance surface, and $\boldsymbol{R}_0$ is the probe position. The expansion of the probe as a phase

aberrated spherical wave is well known,

$$P(\boldsymbol{R}-\boldsymbol{R}_0) = \int_{\text{objective aperture}} \exp\{\mathrm{i}[\boldsymbol{K}\cdot(\boldsymbol{R}-\boldsymbol{R}_0) + \gamma(\boldsymbol{K})]\}\,\mathrm{d}\boldsymbol{K} \quad (1)$$

where $\boldsymbol{K}$ is the transverse component of the incident wavevector χ, and

$$\gamma = \frac{\pi}{\lambda}(\Delta f\theta^2 + \tfrac{1}{2}C_s\theta^4) \quad (2)$$

is the transfer function phase factor for a defocus Δf and spherical aberration coefficient C_s. The scattering angle $\theta = K/\chi$ and $\lambda = 2\pi/\chi$ is the electron wavelength. For a very thin specimen, effects of probe dispersion and absorption may be ignored, and the scattered amplitude ψ_s obtained immediately from the first Born approximation,

$$\psi_s(\boldsymbol{K}_f) = \frac{m}{2\pi\hbar^2}\int \exp[\mathrm{i}\boldsymbol{K}_f\cdot\boldsymbol{R}] \times V(\boldsymbol{R})P(\boldsymbol{R}-\boldsymbol{R}_0)\,\mathrm{d}\boldsymbol{R} \quad (3)$$

where $V(\boldsymbol{R})$ is the projected potential. Integrating the scattered intensity $|\psi_s|^2$ over all final states $\boldsymbol{K}_f$, using the identity

$$\int \exp[-\mathrm{i}\boldsymbol{K}_f\cdot(\boldsymbol{R}-\boldsymbol{R}')]\,\mathrm{d}\boldsymbol{K}_f = (2\pi)^2\delta(\boldsymbol{R}-\boldsymbol{R}') \quad (4)$$

gives the image intensity

$$I(\boldsymbol{R}_0) = \int O(\boldsymbol{R})P^2(\boldsymbol{R}-\boldsymbol{R}_0)\,\mathrm{d}\boldsymbol{R} \quad (5)$$

which represents a convolution of the probe *intensity* profile $P^2(\boldsymbol{R})$ with an object function $O(\boldsymbol{R})$ given by

$$O(\boldsymbol{R}) = \sigma^2 V^2(\boldsymbol{R}) \quad (6)$$

where $\sigma = \chi/2E$ is the interaction constant. Therefore, provided *all* scattered electrons could be collected, we see immediately that incoherent imaging would be obtained with a resolution controlled by the incident probe intensity profile. Note that the Fourier components of the object function are given by

$$\tilde{O}(\boldsymbol{K}) = \int O(\boldsymbol{R})\exp(-\mathrm{i}\boldsymbol{K}\cdot\boldsymbol{R})\,\mathrm{d}\boldsymbol{R} \quad (7)$$

which can be written as a self-convolution of elastic scattering form factors $f(K)$,

$$\tilde{O}(\boldsymbol{K}) = \frac{1}{\chi^2}\int f(\boldsymbol{K}+\boldsymbol{K}')f(\boldsymbol{K}')\,\mathrm{d}\boldsymbol{K}' \quad (8)$$

A similar reciprocal space integral occurs in the object functions for diffuse scattering and inelastic scattering presented later.

Unfortunately, it is not practically possible to detect all scattered electrons, which has led in the past to a widely held but fallacious view that incoherent imaging could not be achieved at high resolution. Over the angular range of the objective aperture the scattered beam cannot be physically distinguished from the unscattered beam. For high-resolution imaging, the objective aperture is necessarily large, so that a significant fraction of the total scattering will inevitably occur within the same angular range. If a hole is cut in the detector to exclude the unscattered beam, some of the scattered beam will also be lost, and the requirements for incoherent imaging cannot be satisfied [5, 6]; this is referred to as the hole-in-the-detector problem.

However, at the cost of reduced signal intensity, this problem may be circumvented to any desired degree of accuracy. Suppose it was possible to collect all the scattered intensity up to some cut-off

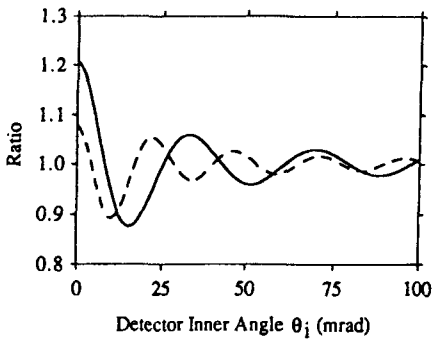


Figure 5. Ratio of detected intensity to the incoherent result for two silicon atoms spaced 1.0 Å (solid line) and 1.5 Å (dashed line) apart for an annular detector with inner angle θ_i.

wavevector K_c. Due to the angular dependence of atomic form factors, with increasing K_c the signal would increase and eventually saturate. If K_c is chosen to collect a large fraction f of the total scattering, sufficient to provide incoherent imaging to our desired accuracy, then collecting all the scattering above K_c will also give an incoherent image, but with the reduced signal level of $1 - f$. In this way, a high angle annular detector can collect a small but *representative* fraction of the total scattering and circumvent the hole-in-the-detector problem. Figure 5 shows how the intensity detected from two silicon atoms approaches the incoherent limit as the inner angle of the annular detector is increased [7]. Significant deviations occur for small inner angles, but at the second minimum in the figure the deviation from incoherent imaging is only ~5%. This gives a useful criterion for the minimum inner detector angle necessary for the incoherent imaging of two objects separated by ΔR,

$$\theta_i = 1.22\lambda/\Delta R \tag{9}$$

The object function is no longer the full potential $V^2(\boldsymbol{R})$, but involves only the sharp part of the potential that generates the high-angle scattering,

$$O(\boldsymbol{R}) = \sigma^2 |V(\boldsymbol{R}) * d(\boldsymbol{R})|^2 \tag{10}$$

Here, $d(\boldsymbol{R})$ is the Fourier transform of the annular detector function $D(\boldsymbol{K})$, which is unity over the detector and zero elsewhere. For incoherent imaging to be a good approximation, $d(R)$ must be significantly narrower than $P(\boldsymbol{R})$. It then localizes contributions to the scattered intensity in a manner similar to the action of the δ function in Eq. (4). The difference, however, is that the detector acts as a spatial frequency filter for the potential $V(\boldsymbol{R})$, selecting only the sharp nuclear contribution. This is the origin of the strong Z sensitivity to the image. It is of course very convenient that the atomic potential does include a wide range of spatial frequencies, since this approach would fail for a specimen comprising a single sine wave potential.

Provided the criterion (9) is met, images of single atoms or small clusters will show no reversals in contrast with lens defocus, and will show enhanced resolution compared to a bright field image, characteristics that were indeed seen experimentally [8]. However, these incoherent characteristics refer only to the transverse plane. Although coherence in the transverse plane is limited to scales smaller than the width of $d(\boldsymbol{R})$, which can be conveniently arranged to be below the intercolumnar spacings we wish to image, for a column of n atoms aligned with the beam direction, the object function becomes

$$O(\boldsymbol{R}) = \sigma^2 |nV(\boldsymbol{R}) * d(\boldsymbol{R})|^2 \tag{11}$$

giving an intensity proportional to the square of the number of atoms in the column. This shows that each atom within

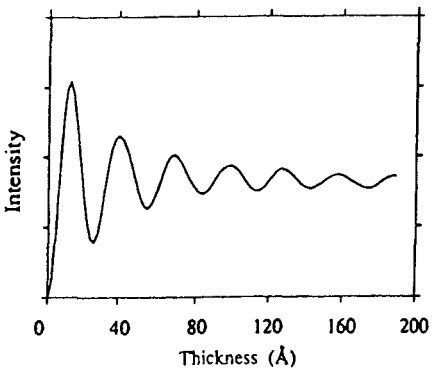


Figure 6. The intensity of zero-layer coherent scattering reaching a 50–150 mrad annular detector calculated kinematically for a 100 kV probe of 10.3 mrad semiangle located over a dumbbell in Si ⟨110⟩.

a particular column contributes coherently to the total intensity scattered by that column. Re-evaluating the experimental data of Isaacson et al. [9] indicates the expected n^2 behavior of intensity on column length [7].

For crystals with thickness t comparable to $2\lambda/\theta_i^2$, the amplitude scattered from the top and bottom of a column will no longer be in phase, and will destructively interfere. This results in the oscillatory thickness dependence, as shown in Fig. 6. Note that in the absence of thermal vibrations the intensity never rises above that of a thin crystal. The reason is clear from a reciprocal space argument; at this thickness it is no longer possible to ignore the curvature of the Ewald sphere, and with increasing column length the scattered intensity becomes more concentrated into the reciprocal lattice spot, thus missing the Ewald sphere. In practice, however, this behavior is masked by the increasing importance of thermal vibrations, which also mask any contribution from HOLZ lines.

2.3.3 Incoherent Imaging with Thermally Scattered Electrons

The need for high scattering angles naturally suggests that thermal diffuse scattering might represent an important, perhaps dominant contribution to the detected intensity. Furthermore, we might suppose that the thermal vibrations themselves could be effective in breaking the coherence of the imaging process, which was the original motivation for the use of a high angle annular detector [10]. We find that thermal vibrations make little difference in the transverse plane; coherence here is very efficiently broken by the detector geometry alone. Thermal vibrations are, however, very effective in breaking the coherence along the columns, dramatically suppressing HOLZ lines and avoiding the oscillatory behavior of coherent zero layer diffraction seen in Fig. 5.

Thermal vibrations are normally treated in the framework of an Einstein model of independently vibrating atoms; in this case the diffuse scattering is by definition generated incoherently. In reality, atomic vibrations are not independent, and positions of near-neighbor atoms will be correlated. In order to examine the effect of these short-range positional correlations on the image intensity we need a phonon model of lattice vibrations. First, however, we consider the Einstein model since it provides explicit atomic cross-sections for the diffuse scattering, although, for the correct high angle behavior, multiphonon scattering processes must be included. We therefore use the approach of Hall and Hirsch [11, 12], which naturally includes all multiphonon processes

by calculating the total time-averaged intensity distribution, and subtracting the Bragg reflections. Integrating this intensity over the Ewald sphere gives directly the Fourier components of the object function,

$$\tilde{O}^{\mathrm{TDS}}(\boldsymbol{K}) = \frac{1}{\chi^2}\int f(|\boldsymbol{s}'|)f(|\boldsymbol{s}-\boldsymbol{s}'|) \times [\exp(-Ms^2) - \exp(-Ms'^2)] \times \exp[-M(\boldsymbol{s}-\boldsymbol{s}')^2]\,\mathrm{d}\boldsymbol{K}' \quad (12)$$

where $s = K/4\pi$ as used in the conventional tabulations [13], and M is the usual Debye–Waller factor.

Now consider the important case of the high-angle annular detector. If the potential is much sharper than the probe intensity profile, we can remove the probe from the integral in Eq. (5), giving

$$I(\boldsymbol{R}_0) \approx P^2(\boldsymbol{R}-\boldsymbol{R}_0)\int O^{\mathrm{TDS}}(\boldsymbol{R})\,\mathrm{d}\boldsymbol{R} = P^2(\boldsymbol{R}-\boldsymbol{R}_0)\tilde{O}^{\mathrm{TDS}}(0) \quad (13)$$

where $\tilde{O}^{\mathrm{TDS}}(0)$ is simply the atomic diffuse scattering distribution integrated over the high-angle detector,

$$\tilde{O}^{\mathrm{TDS}}(0) = \frac{1}{\chi^2}\int_{\text{detector}} f^2(|\boldsymbol{s}'|) \times [1-\exp(-2Ms'^2)]\,\mathrm{d}\boldsymbol{K}' = \sigma_{\mathrm{at}} \quad (14)$$

the atomic cross-section for diffuse scattering into the annular detector. For sufficiently high angles the term in square brackets tends to unity, and the cross-section becomes the total elastic scattering intensity integrated over the detector. For a column of n atoms aligned with the beam direction, these cross-sections are summed, giving an intensity which scales as n in the limit of a thin crystal where probe dispersion and absorption are negligible,

$$O(\boldsymbol{R}) = n\sigma_{\mathrm{at}}\delta(\boldsymbol{R}) \quad (15)$$

We now consider the effects of correlated vibrations on the diffuse scattering. In the context of a high-angle annular detector, transverse incoherence is ensured by the detector geometry, so again we need only consider correlations for a column of atoms aligned with the beam direction. The important physical concept to emerge is the existence of a finite coherence *length* l_c along a column. Crystals of thickness $t \ll l_c$ will show the n^2 dependence characteristic of coherent scattering, as seen in Eq. (11). In crystals with $t \gg l_c$, the coherent component will be small, and columnar intensities will show the linear behavior predicted by the Einstein model of diffuse scattering. Note, however, that l_c cannot be obtained from the Einstein model of independently vibrating atoms; here the coherence length is by definition zero for the diffuse component, and t for the coherent component. In reality, the degree of correlation is highest for near neighbors, and reduces smoothly with increasing separation.

To examine the role of residual correlations it is necessary to consider phonons explicitly. Consider a column of atoms with instantaneous deviations u_i from their lattice positions $\boldsymbol{r}_i = (\boldsymbol{R}_i, z_i)$ scattering to a high-angle detector, so that all scattering occurs in close proximity to the atomic sites, and transverse incoherence is ensured. An object function can then be written including all zero layer, HOLZ, and diffuse scattering

as [14]

$$I^{H}(\boldsymbol{R}_0) = \frac{1}{\chi^2}\sum_{m,n}\int_{\text{detector}} f^2(\boldsymbol{s}) \times \exp\{i\boldsymbol{k}\cdot(\boldsymbol{r}_m - \boldsymbol{r}_n)\} \times \langle \exp\{i\boldsymbol{k}\cdot(\boldsymbol{u}_m - \boldsymbol{u}_n)\}\rangle \, d\boldsymbol{K} \quad (16)$$

where $\boldsymbol{k} \approx (\boldsymbol{K}, K^2/2\chi)$ is the scattering vector and the angled brackets denote the time average. A treatment due to Warren [15] incorporates phonons by assuming a Debye dispersion relation, equipartition, and replacing the Brillouin zone boundary at q_B by a sphere of equal volume, whereupon the time average in Eq. (16) reduces to an analytical form,

$$W_{mn} = \langle \exp\{i\boldsymbol{k}\cdot(\boldsymbol{u}_m - \boldsymbol{u}_n)\}\rangle = \exp\left[2M\left(\frac{\mathrm{Si}(2\pi q_B r_{mn})}{2\pi q_B r_{mn}} - 1\right)\right] \quad (17)$$

This factor expresses the degree of correlation between the mth and nth atoms in a column, where $r_{mn} = |\boldsymbol{r}_m - \boldsymbol{r}_n|$ and $\mathrm{Si}(x)$ is the sine integral function

$$\mathrm{Si}(x) = \int_0^x \frac{\sin u}{u}\,du \quad (18)$$

The partial coherence between near neighbors in the column is shown in Fig. 7, where W_{0n} is seen to fall smoothly from 1 at $m = 0$ to e^{-2M} for large m. If we use an Einstein dispersion relation, with constant frequency for all modes and polarizations, we recover the Einstein model, with a correlation factor given by

$$W_{0n} = \exp[2M(\delta_{0n} - 1)] \quad (19)$$

This is again 1 at $m = 0$, but now takes the value e^{-2M} for all other atoms, that is, near-neighbor correlations are specifically excluded.

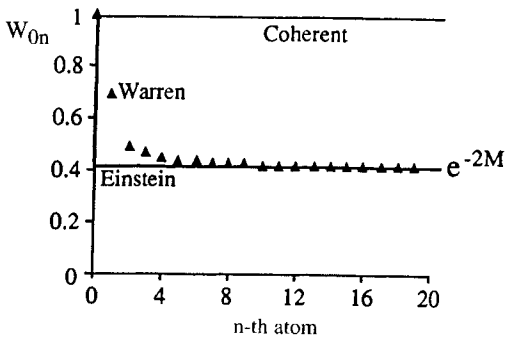


Figure 7. Correlation factor between atoms separated by n spacings along a column calculated in the Warren approximation (triangles), the Einstein limit and the coherent limit for $M = 0.45$ and $\theta_i = 75$ mrad. The arrowed vertical line defines a packet within which partial coherence exists.

Physically, we can picture the columnar coherence volume as defining a packet surrounding each atom, within which partial coherence exists, but outside of which the atoms appear uncorrelated. Coherent interference effects are observed in the form of thickness oscillations for columns shorter than the length of the packet, while for longer columns the thickness dependence becomes linear, but with a slope that can be above or below the Einstein value depending on whether the interference effects within each packet are constructive or destructive in nature. Figure 8 shows the thickness dependence for a column of silicon atoms, showing the initial t^2 dependence, some destructive interference, and the eventual linear behavior. Shown also are the coherent and diffuse components calculated in the Einstein model. In this case it can be seen that the coherence effects within a packet are constructive. Generally, however, the changes in columnar cross-section due to correlation effects are relatively small for high-angle detectors, but can be significant with low-angle detectors. This effect could be utilized to

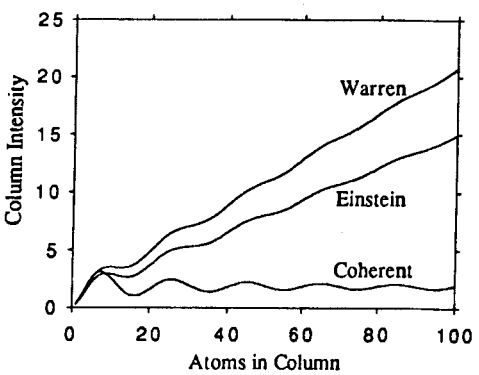


Figure 8. Calculated thickness dependence of elastic scattering from a single column of silicon atoms calculated with a phonon model for diffuse scattering due to Warren [15], compared to the coherent and diffuse components calculated in the Einstein model (inner detector angle 50 mrad, atom separation 1.5 Å, Debye–Waller parameter $M = 0.45$).

boost signal levels from beam-sensitive specimens or small clusters by tuning the packet length to match the specimen thickness.

2.3.4 Incoherent Imaging using Inelastically Scattered Electrons

The optical equation for incoherent imaging with inelastically scattered electrons was first derived by Rose [16], and expressed in a form explicit to STEM by Ritchie and Howie [17]. The derivation is somewhat more elaborate than the simple case of elastic scattering given above, but follows similar lines. We give here Ritchie and Howie's equation (11) for the image intensity when all scattered electrons are collected, which can again be written as a convolution of the probe intensity profile with an inelastic object function $O'(\boldsymbol{R})$,

$$I'(\boldsymbol{R}_0) = \int P^2(\boldsymbol{R} - \boldsymbol{R}_0) O'(\boldsymbol{R})\, \mathrm{d}\boldsymbol{R} \qquad (20)$$

where

$$O'(\boldsymbol{R}) = \left(\frac{e^2}{\pi\hbar \mathrm{v}}\right)^2 \times \left| \int \frac{\rho_{n0}}{q^2}(\boldsymbol{q}) \exp(-\mathrm{i}\boldsymbol{K}\cdot\boldsymbol{R})\, \mathrm{d}\boldsymbol{K} \right|^2 \qquad (21)$$

Here we write the momentum transfer $\boldsymbol{q} = (\boldsymbol{K}, q_z)$ to recognize the minimum momentum transfer at zero scattering angle, and introduce the matrix elements ρ_{n0} for the transition from an initial state $|0\rangle$ to a final state $|n\rangle$. Ritchie and Howie showed further that the matrix elements in Eq. (21) may be calculated in the semiclassical approximation, that is, assuming no deflection of the fast electron trajectory, which was convenient for their study of the excitation of surface plasmons (see also the discussion by Batson [18]).

Again, the object function can be expressed as a form factor in reciprocal space [19],

$$\tilde{O}'(\boldsymbol{K}) = \left(\frac{2e^2}{\hbar \mathrm{v}}\right)^2 \int \frac{\rho_{n0}(\boldsymbol{q}')}{q'^2} \frac{\rho^*_{n0}(\boldsymbol{q}' + \boldsymbol{K})}{|\boldsymbol{q}' + \boldsymbol{K}|^2} \mathrm{d}\boldsymbol{K}' \qquad (22)$$

where

$$O'(\boldsymbol{R}) = \frac{1}{(2\pi)^2} \int \tilde{O}'(\boldsymbol{K}) \exp(\mathrm{i}\boldsymbol{K}\cdot\boldsymbol{R})\, \mathrm{d}\boldsymbol{K} \qquad (23)$$

For atomic resolution analysis we are primarily concerned with inner shell excitations, and here a hydrogenic model developed by Maslen and Rossouw is particularly useful [20–24]. Now, an inelastic excitation with scattering vector $\boldsymbol{q}$ for the fast electron is modeled by excitation of the K-shell electron to the

continuum of the hydrogenic atom, representing ejection of a secondary electron with wavevector κ. As the secondary is not observed experimentally, the process is integrated over all secondary electron emission directions, and the matrix product $\rho_{n0}(\boldsymbol{q}')\rho_{n0}(\boldsymbol{q}'+\boldsymbol{K})$ is replaced with

$$\int \rho_{n0}(\boldsymbol{q}', \kappa)\rho^*_{n0}(\boldsymbol{q}'+\boldsymbol{K}, \kappa)\,\mathrm{d}\kappa \qquad (24)$$

Analytical expressions have been given for these integrals [25, 26], and their general properties have been considered by Kohl and Rose [27].

Figure 9 plots the full-width half-maxima of object functions calculated for a number of elements, assuming that all the scattered electrons are collected. Somewhat surprisingly it would appear that atomic resolution is possible even for the lightest elements. This is a significant difference from impact parameter estimates based on the angular widths of scattering distributions [28–30], and reflects the strong Z dependence of the hydrogenic model. Furthermore, the quantum mechanical expression is insensitive to the beam energy, predicting object function widths that are practically identical for 100 and 300 kV electrons.

Although analytical expressions are available for L- and M-shell cross-sections, none exist for the Fourier components $\tilde{O}'(\boldsymbol{K})$. However, if we are only concerned with excitations near threshold it is perfectly reasonable to model the process as bound-to-bound transitions within the hydrogenic atom, as recently suggested by Holbrook and Bird [31]. They find widths for L-shell excitations that are very comparable to those for K-shell excitation at the same energy, findings that are consistent with recent experimental results. Batson [32] has found changes in the silicon L fine structure on moving the probe one or two atomic spacings, while Browning et al. [33] obtained atomic resolution at the cobalt L edge, using a $CoSi_2$–Si (111) epitaxial interface as an edge resolution

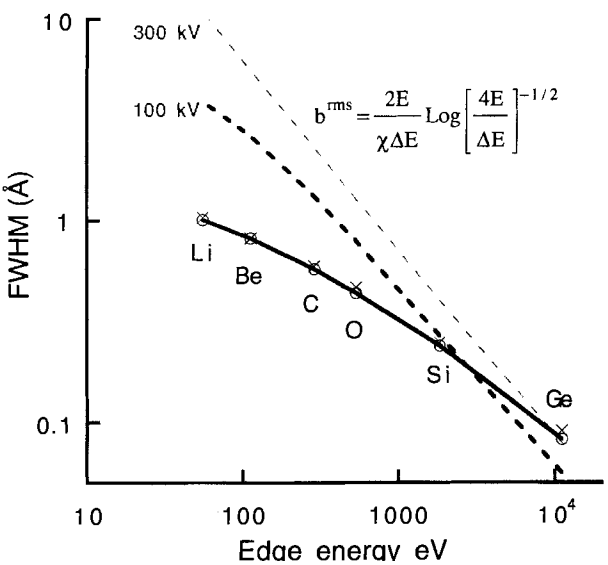


Figure 9. Full-width half-maxima of object functions for K-shell excitation of various elements calculated with a hydrogenic model for a 50 eV ejected secondary assuming all scattered electrons are collected. Circles are for 100 kV incident electrons, crosses for 300 kV. Dashed lines show impact parameters estimated from the angular widths of scattering distributions [30].

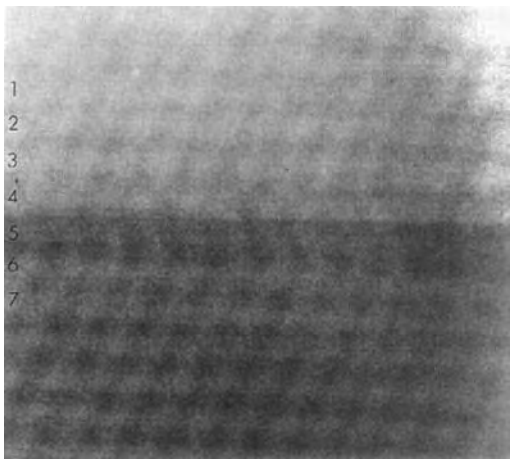


Figure 10. The Z-contrast image of a $CoSi_2$–Si (111) epitaxial interface used as an edge resolution test specimens for EELS. Cobalt atoms image bright; planes numbered correspond to the spectra in Fig. 11.

test specimen. Figure 10 is an image of this interface taken with a 100 kV VG Microscopes HB501UX, which reveals a structure not previously considered for this interface [34]. The last plane of the silicon is seen to be in a rotated configuration, which maintains the eightfold coordination of the cobalt. EELS spectra recorded from each plane in turn clearly demonstrated the atomic resolution capability. In Fig. 11 the cobalt L edge is seen to drop from 86 to 7% of its bulk value in moving a single plane across the interface, a distance of 2.7 Å.

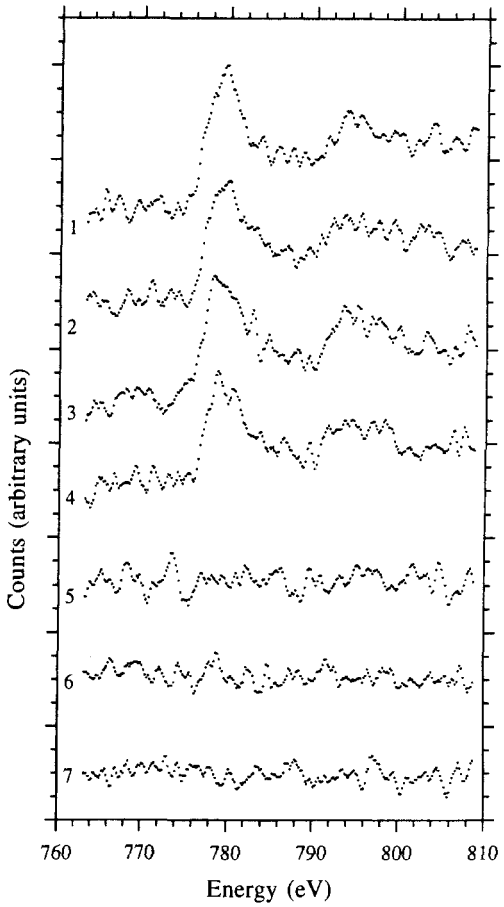


Figure 11. EELS spectra obtained phase by plane across the interface of Fig. 10 showing atomic resolution at the cobalt L edge.

2.3.5 Probe Channeling

In practice, specimens for which the Born approximation is valid are rarely, if ever, encountered. Even a single heavy atom is sufficient to invalidate the Born approximation, and for a typical microscope specimen one might expect dynamical scattering to complicate the situation and necessitate a full image simulation, a return to a reliance on model structures. However, experimentally it is found that the characteristics of incoherent imaging are preserved to large specimen thicknesses [35–38]. In fact, with a 300 kV scanning transmission electron microscope, structure imaging remains to quite remarkable thicknesses.

The reason for this lies in the channeling behavior of a coherent STEM probe, in

particular, the effect of the superposition required to describe the convergent probe. A plane wave entering a crystal forms the set of Bloch states,

$$\psi(\boldsymbol{R},z)=\sum_j \epsilon^j(\boldsymbol{K})\tau^j(\boldsymbol{K},\boldsymbol{R}) \times \exp(-\mathrm{i}s^j(\boldsymbol{K})z/2\chi) \times \exp(-\mu^j(\boldsymbol{K})z) \tag{25}$$

where $\tau^j(\boldsymbol{K},\boldsymbol{R}) = b^j(\boldsymbol{K},\boldsymbol{R})\exp(\mathrm{i}\boldsymbol{K}\cdot\boldsymbol{R})$ are two-dimensional Bloch states of transverse energy $s^j(\boldsymbol{K})$, absorption $\mu^j(\boldsymbol{K})$ and excitation coefficients $\epsilon^j(\boldsymbol{K})$. The states of greatest transverse energy take on the form of molecular orbitals about the atomic strings, as seen in Fig. 12. They propagate with slightly different wavevectors through the crystal thickness, and it is their interference that leads to the familiar dynamical diffraction effects. Now to form a probe we make a coherent superposition of phase-aberrated plane waves over the objective aperture (see Eq. (1)), so that the wave function inside the crystal becomes

$$\psi(\boldsymbol{R}-\boldsymbol{R}_0,z) = \sum_j \int_{\text{objective aperture}} \epsilon^j(\boldsymbol{K})b^j(\boldsymbol{K},\boldsymbol{R}) \times \exp(-\mathrm{i}s^j(\boldsymbol{K})z/2\chi)\exp(-\mu^j(\boldsymbol{K})z) \times \exp\{\mathrm{i}[\boldsymbol{K}\cdot(\boldsymbol{R}-\boldsymbol{R}_0)+\gamma(\boldsymbol{K})]\}\,\mathrm{d}\boldsymbol{K} \tag{26}$$

The effect of this angular integration depends very strongly on the nature of the particular Bloch state. Tightly bound 1s states overlap little between neighboring

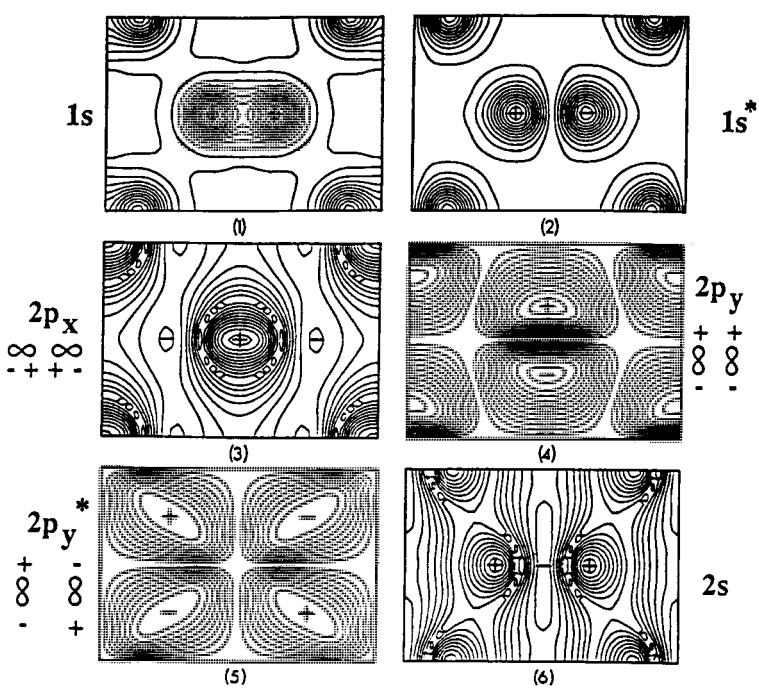


Figure 12. Bloch states for a 100 kV electron in silicon ⟨110⟩, with their molecular orbital assignments.

strings and are therefore nondispersive. They add in phase during the angular integration. In contrast, states such as 2s and 2p are significantly more extended and do overlap with states on neighboring columns. They then become dispersive and interfere destructively when forming the superposition. Calculations confirm that the intensity at the atomic columns is dominated by the tightly bound s-type states [35–37]. This explains how an incident probe propagating through the crystal becomes compacted around the columns themselves, taking on the character of the s states [39, 40]. It is in this way that we obtain a rather simple columnar channeling behavior.

For signals that originate from close to the atomic sites, such as the high angle elastic and diffuse scattering, a good approximation to the full wavefunction of Eq. (26) is to include only s states in the Bloch wave sum. Since they are nondispersive, they may be taken outside the angular integration, giving

$$\begin{aligned}\psi(\boldsymbol{R}-\boldsymbol{R}_0,z)&\approx b^{1s}(\boldsymbol{R},0)\exp(-is^{1s}(0)z/2\chi)\\&\quad\times\exp(-\mu^{1s}(0)z)\\&\quad\times\int_{\text{objective aperture}}\frac{\epsilon^{1s}(\boldsymbol{K})}{\epsilon^{1s}(0)}\\&\quad\times\exp\{i[\boldsymbol{K}\cdot(\boldsymbol{R}-\boldsymbol{R}_0)+\gamma(\boldsymbol{K})]\}d\boldsymbol{K}\end{aligned}\tag{27}$$

We once more recover the optical equation for incoherent imaging,

$$I(\boldsymbol{R},t)=O(\boldsymbol{R},t)*P^2_{\text{eff}}(\boldsymbol{R})\tag{28}$$

where $O(\boldsymbol{R},t)$ is now the thickness-dependent object function,

$$\begin{aligned}O(\boldsymbol{R},t)&=\epsilon^{1s^2}(0)b^{1s^2}(\boldsymbol{R},0)\\&\quad\times\left(\frac{1-\exp(-2\mu^{1s}(0)t)}{2\mu^{1s}(0)}\right)O^{B}(\boldsymbol{R})\end{aligned}\tag{29}$$

in which the appropriate Born approximation object function $O^{B}(\boldsymbol{R})=O(\boldsymbol{R})$ or $O^{TDS}(\boldsymbol{R})$ is scaled by the columnar channeling effect, and

$$\begin{aligned}P^2_{\text{eff}}(\boldsymbol{R})=\Bigg|\frac{1}{\epsilon^{1s}(0)}&\int_{\text{objective aperture}}\epsilon^{1s}(\boldsymbol{K})\\&\times\exp\{i[\boldsymbol{K}\cdot(\boldsymbol{R}-\boldsymbol{R}_0)\\&+\gamma(\boldsymbol{K})]\}\,d\boldsymbol{K}\Bigg|^2\end{aligned}\tag{30}$$

is an effective probe intensity profile which includes the angular fall-off in s-state excitation. This fall-off is quite small over the range of a typical objective aperture, and only broadens the probe by around 10%. The same situation will hold for highly localized inelastic scattering, and may also hold approximately for less local object functions, although the degree of s-state dominance away from the atom sites has yet to be investigated.

This formulation models the imaging process as weak scattering out of the s-state wavefield. Note that since the s states do not broaden with depth into the crystal, the assumption of a nondispersive probe used in the Born approximation derivations is not violated. The thickness dependence of the object function is shown in Fig. 13 for silicon and germanium in the ⟨110⟩ projection. As the s states are peaked at the atomic sites, they are absorbed quite rapidly, especially with high-Z columns as

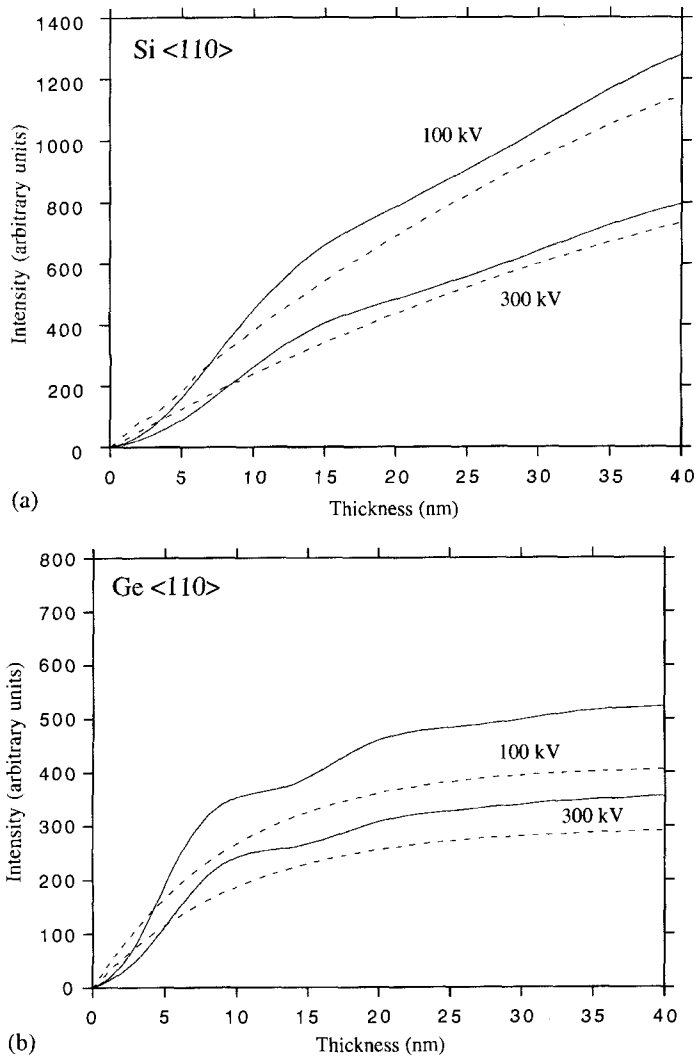


Figure 13. Thickness dependence of the object function in (a) silicon ⟨110⟩ and (b) germanium ⟨110⟩ calculated using s states alone (dashed lines) compared to the full dynamical calculation (solid lines). Calculations assume an Einstein absorption model.

in the case of germanium. This leads to reduced contrast between heavy and light columns with increasing thickness. Also shown in Fig. 13 is the object function calculated with all Bloch states included, in which residual dynamical effects are seen at small thicknesses. A more accurate model for the object function [36] is to include the interference of the s states with the incident probe, which gives a better fit to the full dynamical curve.

This channeling behavior explains how incoherent imaging characteristics are maintained in thick crystals, and how weak scattering models can be simply modified to predict elastic or inelastic image behavior, even in the presence of the dynamical diffraction. An important additional benefit of imaging with only the highly localized Bloch states is that the object function remains highly local. This means that the columnar scattering

power is independent of the type and arrangement of surrounding atomic columns. Image interpretation remains highly intuitive, in contrast to coherent imaging in which interference with nonlocal interface states may complicate interpretation.

2.3.6 Applications to Materials Research

2.3.6.1 Semiconductors

One of the alluring properties of a 300 kV scanning transmission electron microscope is a demonstrated probe size of 1.26 Å, less than the separation of the 'dumbbells' seen in the ⟨110⟩ projection of diamond cubic semiconductors. In compound semiconductors, the two columns of the dumbbell are different species; the incoherence of the image, coupled with its Z contrast, therefore enables the sublattice polarity to be determined directly from the Z-contrast image [41].

Figure 14 shows GaAs imaged in the ⟨110⟩ projection with the VG Microscopes HB603U scanning transmission electron microscope. Although gallium and arsenic are only two atomic numbers apart in the periodic table, their scattering cross-sections differ by approximately 10%, depending upon the inner detector angle. Here, with an inner angle of 30 mrad, they are distinguishable in the raw image; the difference in scattering power is seen from the line trace below. This direct structure image is maintained up to thicknesses of the order of 1000 Å.

A structure image of this nature is a convolution of the probe intensity profile with the object; it should therefore be possible to invert the image directly by deconvolution. Unfortunately, it is well known that deconvolution procedures tend to enhance high-frequency noise, so we have employed a maximum entropy method which has an outstanding noise rejection capability [42]. The maximum entropy image is a blank page, that is, a uniform array of intensity with no information content. This is of course a poor fit to any actual image, but is used as a starting point in the maximum entropy iteration, which proceeds to find an image of constrained maximum entropy, which does provide a good fit to the image data. The probe intensity profile is

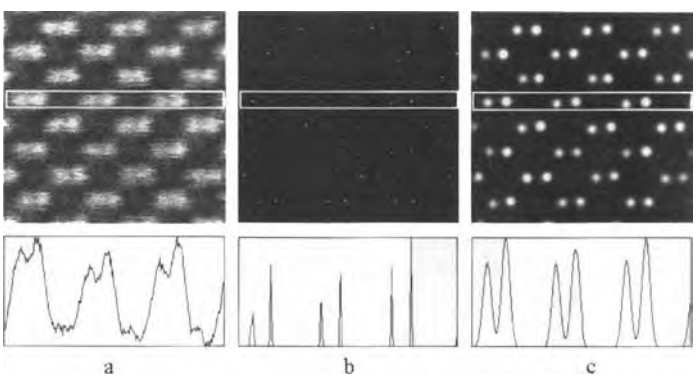


Figure 14. Images of GaAs ⟨110⟩ in the HB603U 300 kV scanning transmission electron microscope. (a) Raw image, with arsenic columns showing the expected ~10% higher scattering power, (b) a maximum entropy reconstruction of the object, and (c) a reconstructed image. Line traces below each image show the vertically averaged intensity within the rectangles outlined.

assumed as an input parameter, though the inversion is not critically dependent on the form assumed. From the line trace in Fig. 14a, it is a simple matter to estimate the probe width, and the maximum entropy object for a Lorentzian probe of this size is shown in Fig. 14b. The reconstructed positions of all gallium and arsenic columns are each within 0.1 Å of their actual positions, and their relative intensity is as expected. This information is appreciated more easily by reconvoluting the object function with a small Gaussian, to give the reconstructed image of Fig. 14c. Notice how the raw image contains a secondary maximum in the channels between the dumbbells, arising from the tails of the probe intersecting the six surrounding dumbbells. These features are not part of the object, and are successfully removed from the reconstructed image.

The greatest value of this direct inversion capability is that it can reveal unexpected features of the object that would not have been included in any model structure. Figure 15 shows an image of a perfect edge dislocation core at a CdTe(001)–GaAs(001) interface reconstructed in this way. It is clearly inconsistent with the five- and seven-membered ring structure of the Hornstra core [43], which is often assumed for these materials, since it shows a fourfold ring surrounded by distorted sixfold rings [41]. Such data can be used as a starting point for theoretical studies of dislocation properties.

Not all interfaces contain surprises, as demonstrated by Fig. 16, a section of a 39° ⟨110⟩ symmetric tilt boundary in silicon ($\Sigma = 9, \{221\}\langle 110\rangle$), viewed along the [110] direction [44]. The boundary is seen to consist of a periodic array of perfect edge dislocations with their line direction

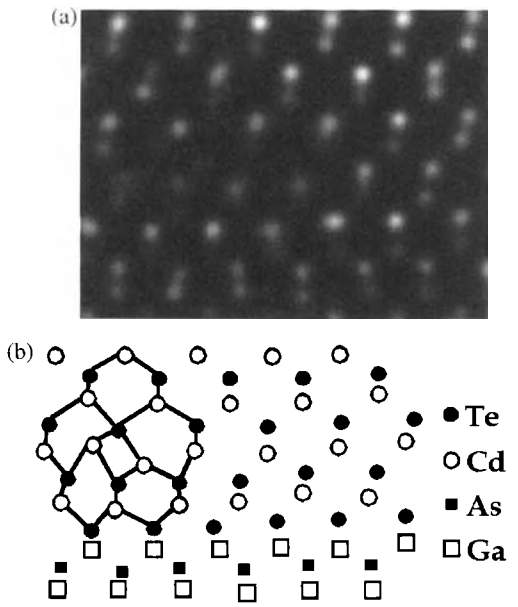


Figure 15. A new core structure observed for an edge dislocation at a CdTe (001)–GaAs (001) interface. The core comprises four- and sixfold rings rather than the five- and sevenfold rings of the Hornstra structure.

parallel to the ⟨110⟩ tilt axis. This is seen clearly to comprise the five- and seven-membered ring structure first shown by Krivanek et al. [45]. In contrast, the symmetric $\Sigma = 13, \{150\}\langle 001\rangle$, tilt boundary in silicon (Fig. 17) does show an unexpected arrangement of dislocation cores [46]. Here the boundary is a contiguous array of six dislocations, consisting of a perfect edge dislocation (labeled 1) and two perfect mixed dislocations arranged as a dipole (labeled 2 and 3), followed by the same three cores mirrored across the boundary plane (labeled 1′, 2′, and 3′). These cores are connected, but remain individually distinct, and tetrahedral bonding is retained through a combination of five-, six-, and seven-membered rings. This structure differs from the aperiodic

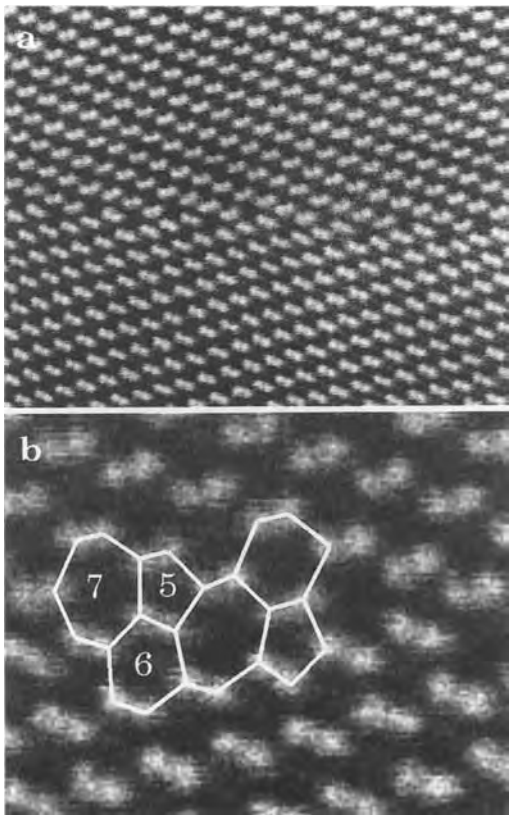


Figure 16. $\Sigma = 9, \{221\}\langle 110\rangle$, symmetric tilt boundary in silicon (viewed along the [110] direction), showing its five- and seven-membered ring structure.

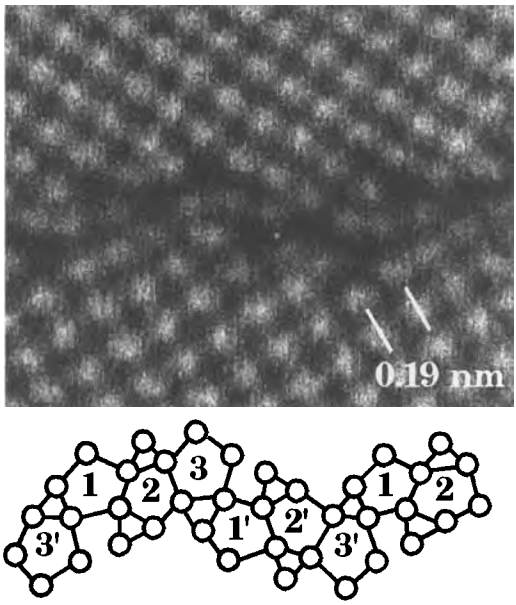


Figure 17. $\Sigma = 13, \{150\}\langle 001\rangle$, symmetric til ary in silicon showing an unexpected arrang dislocation cores comprising five-, six-, an membered rings.

boundary containing multiple structures reported previously [47], and is not the structure predicted from the principle of continuity of boundary structure [48].

Structure images of this nature are greatly complemented by the ability to perform EELS at atomic resolution. Given our present understanding of the intrinsic width of inelastic object functions, atomic resolution is to be expected for all inner shells likely to be used in microanalysis, whether in a 100 kV or a 300 kV STEM instrument. A spectacular demonstration of the power of combined incoherent structure imaging and EELS is shown by the image of a CdTe Si(100) interface in Fig. 18. Gro MBE, the image has been high-p tered to remove the high backgroun

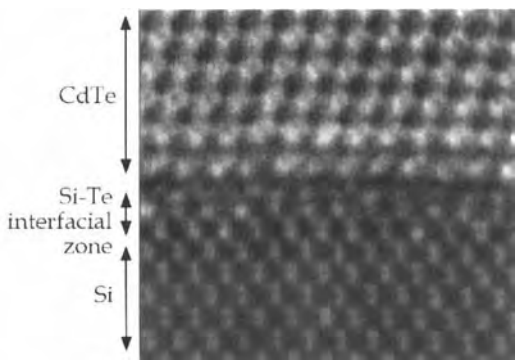


Figure 18. Complex atomic structure at (111)–Si (100) interface grown by MBE. ends on the tellurium sublattice, but EELS the presence of tellurium diffused in the monolayers of the silicon substrate. Tellur columns are seen bright.

the film, and the lattice polarity and interfacial structure are strikingly revealed. The film terminates at tellurium, as was expected, but the two monolayers below do not image in the positions expected for silicon. EELS reveals that these planes contain a substantial fraction of tellurium; in fact, certain columns are seen to be far brighter than can be accounted for by image noise, and must therefore be identified as tellurium-rich. That these effects are a real part of the material is confirmed by the fact that the cadmium EELS signal did indeed show atomic abruptness at the interface. Insights of this nature show directly how the growth procedure determines the interface structure. In turn, the interface structure controls the nucleation of extended defects, such as twins and dislocations, which can then propagate through the entire film and will dominate its electrical properties. With *Z*-contrast structure imaging and atomic resolution EELS, interface engineering becomes a real possibility.

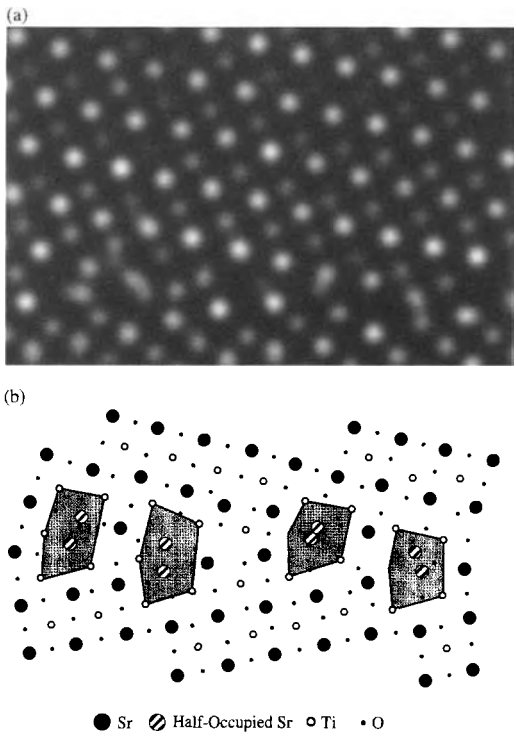


Figure 19. (a) Reconstructed image and (b) atomic structure for a $\Sigma = 85, \{920\}\langle 001\rangle$ symmetric tilt boundary in $SrTiO_3$ deduced using the 100 kV STEM. Strontium columns (seen bright) are spaced 3.9 Å apart in the bulk crystal.

2.3.6.2 Ceramics

An early application of combined *Z*-contrast imaging and atomic resolution EELS was to [001] tilt grain boundaries in $SrTiO_3$ [49]. The strontium and titanium sublattices could be distinguished clearly even with the 100 kV STEM instrument, and EELS spectra could be taken from the boundary plane to compare with spectra from the bulk. In fact, the fine structure changed only a little, showing that the titanium at the boundary remained octahedrally coordinated to oxygen, though in a somewhat distorted configuration. The structure deduced for a {920} grain boundary is shown in Fig. 19, and revealed a number of interesting features. Most significantly, the columns shown hatched are located on the strontium sublattice, but clearly cannot both be present at such a small separation. The simplest model consistent with the experimental data is that these columns are half-filled, that is, sites exist in both halves of the bicrystal, but cannot be simultaneously occupied due to like-ion repulsion. Half-occupancy overcomes the problem, and gives two columns in projection consistent with the image. Alternatively, they can be considered as a single reconstructed strontium column.

Such sites give insight into how impurities may segregate into the boundary plane; they have significant excess volume and could be attractive sites for high atomic radius impurity species.

These two half-occupied strontium columns form the core of a grain boundary dislocation with Burgers vector ⟨100⟩. This Burgers vector corresponds to a single lattice plane in the bulk, that is, a ⟨100⟩ plane can terminate from one side of the boundary or the other, but not both simultaneously. This causes the structure to be microscopically asymmetric, even though the boundary is macroscopically symmetric, a form of microfacetting. The reasons for it can be understood from simple elasticity arguments [50].

Examining 36° and 67° symmetric tilt boundaries revealed further structural units, again all characterized by the presence of half-filled columns. For these boundaries, the structural units terminated lattice planes from each side of the boundary, preserving the microscopic symmetry of the atomic structure. The set of structural units assigned originally [50] allowed structure models to be constructed for any symmetric [001] tilt grain boundary using the principle of continuity of boundary structure. However, if we regard unit cells and fragments thereof as simple spacer units (Fig. 20), it is possible to account for all the observed structures using the three pentagonal cores B, $B^{\#}$ and C. The original {920} boundary is given by the sequence $AB^{(1)}AB^{2}A\ldots$ where the labels 1 and 2 represent {100} planes terminating from different halves of the bicrystal. With A^{*} the 36° boundary contains the same B units, as well as additional units of the same geometry but on the other sublattice, that is, a core containing titanium half-columns ($B^{\#}$ in

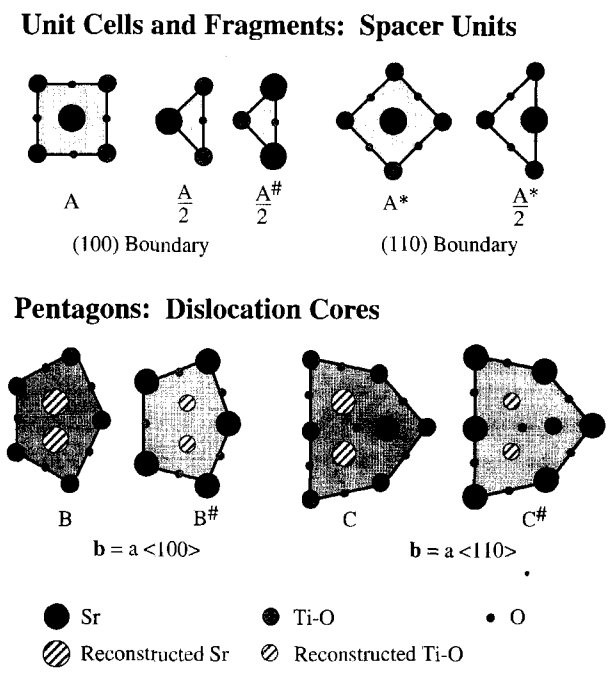


Figure 20. Structural units for all symmetric and asymmetric ⟨001⟩ tilt grain boundaries in $SrTiO_3$.

Fig. 20). Thus, only a single type of lattice dislocation, in these two variants, is needed to account for all symmetric boundaries. In this description, continuity of boundary structure is no longer maintained but the number of units required is reduced.

One advantage of this structural unit description is that the same core can be used in its asymmetric configuration to build asymmetric boundaries. Two different units are required, as it is known from crystal lattice geometry that two types of dislocations are needed to describe an asymmetric boundary. With the 67° grain boundary, cores with a ⟨110⟩ Burgers vector are used (C in Fig. 20) in the sequence $A^{\#}C\ldots$. Images of a 45° symmetric bicrystal [51] which had decomposed into a set of small asymmetric facets are shown in Fig. 21. Units with both ⟨100⟩ and ⟨110⟩ Burgers vectors are used, but the ⟨110⟩ unit now has a core containing Ti half columns ($C^{\#}$ in Fig. 20). In these asymmetric boundaries, the same two units coexist over the entire misorientation range, with different frequencies relative to the spacer unit cells.

Such insights into the structure of grain boundaries and likely impurity sites should enable the important link to be established between grain boundary structure and properties. A recent spectacular advance in this area has come from images of $YBa_2Cu_3O_{7-x}$ grain boundaries, grown on $SrTiO_3$ bicrystals. It might be expected that due to their similar crystal structures the same units exist in $YBa_2Cu_3O_{7-x}$ grain boundaries as in $SrTiO_3$, and this is confirmed by Fig. 22. It is interesting to note that almost the entire grain boundary is asymmetric, waving back and forth about the orientation defined by the substrate.

To gain insight into the superconducting properties of such boundaries we have used a bond valence sum analysis to examine the region around the core where the copper valence differs from its bulk value necessary for superconductivity. The fact that the asymmetric cores contain partially occupied copper columns would be expected to have a significant effect on the copper valence, and this is borne out by the calculations. A substantial nonsuperconducting zone exists around each structural unit. For a low-angle boundary this leads to a rapid reduction in critical current for small orientation angles [52]. At about 10°, the dislocation strain fields overlap sufficiently to cause the boundary to appear as a 'wall' of nonsuperconducting material, through which the supercurrent must tunnel. Now we can understand the puzzling results in the literature reporting the exponential reduction of critical current with misorientation θ [53, 54]. Unit C (see Fig. 20) has the larger Burgers vector, the most strain and the largest nonsuperconducting zone associated with it. Its frequency in the boundary plane increases as $(1 - \cos\theta)$, and the width of the nonsuperconducting zone inferred from bond valence sum analysis increases roughly linearly with misorientation. The exponential decrease in critical current is therefore naturally explained [55].

The implications of such studies for improving transport characteristics are clear. Attempts to improve the high-angle behavior should focus on engineering unit C (see Fig. 20) to 'repair' the copper valence; for applications to wires, efforts should be focused on unit $B^{\#}$ since this is the dominant dislocation at small misorientation angles. *Z*-contrast structure

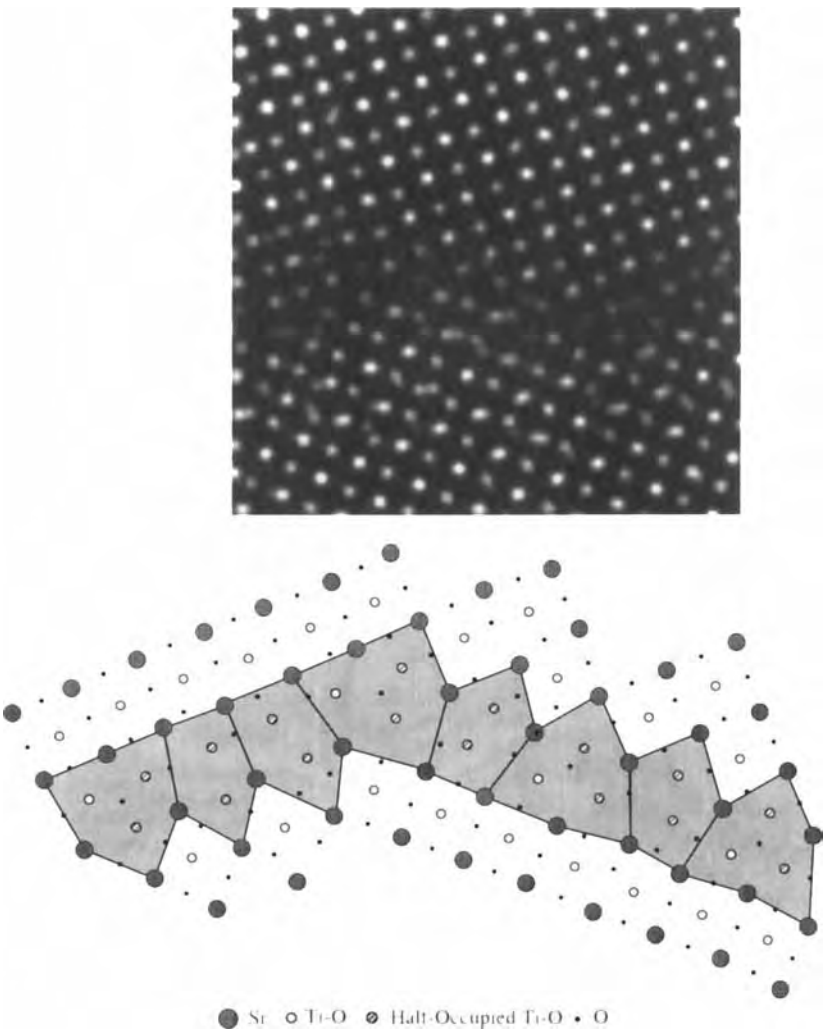


Figure 21. Reconstructed 100 kV image of a macroscopically symmetric 45° tilt boundary in $SrTiO_3$: decomposition into facets, with deduced atomic structure.

imaging with spectroscopy can form an invaluable tool for understanding grain boundary structure–property relationships, as a route to grain boundary engineering. In the world of high-T_c materials, where grain boundaries are necessary to distinguish competing mechanisms [56], these techniques would seem to have particular potential.

2.3.6.3 Nanocrystalline Mate

Of increasing technological impo nanocrystalline materials re another area where Z-contrast m offer major advantages. It is the co tion of freedom from coherent inter effects, such as Fresnel fringes speckle pattern from an amo

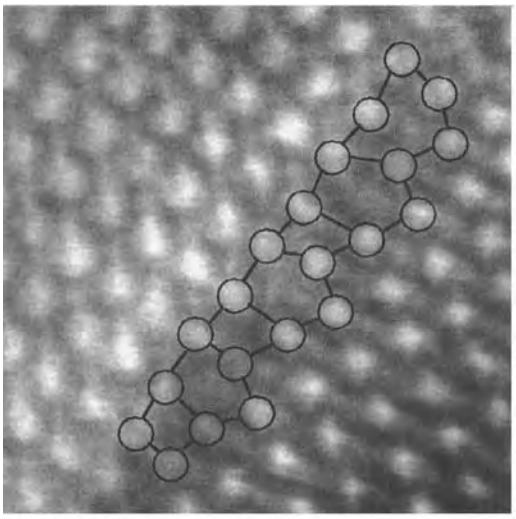


Figure 22. Z-contrast image (300 kV) of a predominantly asymmetric tilt boundary in $YBa_2Cu_3O_{7-x}$, grown on a symmetric 30° $SrTiO_3$ bicrystal. Structural units are equivalent to units $B^{\#}$ and $C^{\#}$ for $SrTiO_3$ (see Fig. 20).

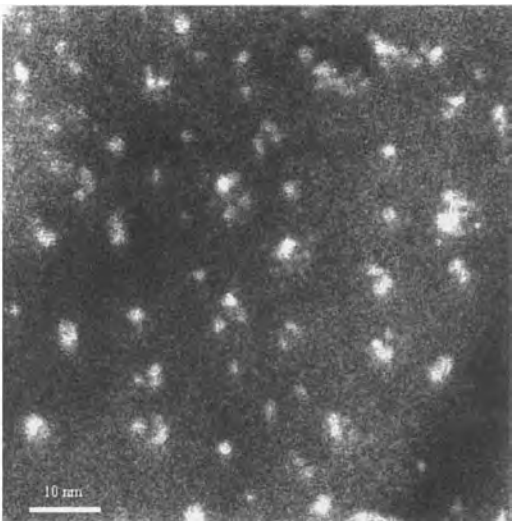


Figure 23. Z-contrast images (300 kV) of silicon nanocrystals formed by implantation into SiO_2 revealing nanometer-sized clusters.

support, with the Z-contrast that provides the insight. As an example, Fig. 23 shows images of nanocrystalline silicon, grown in amorphous SiO_2 by an ion implantation and annealing procedure [57]. Although the Z-contrast between SiO_2 and silicon is not great, because it is the only source of image contrast, the small particles are clearly visible. Surprisingly, they were found to have an internal structure, appearing as clusters of nanometer-sized particles. The smallest of these clusters can contain at most a few tens of atoms.

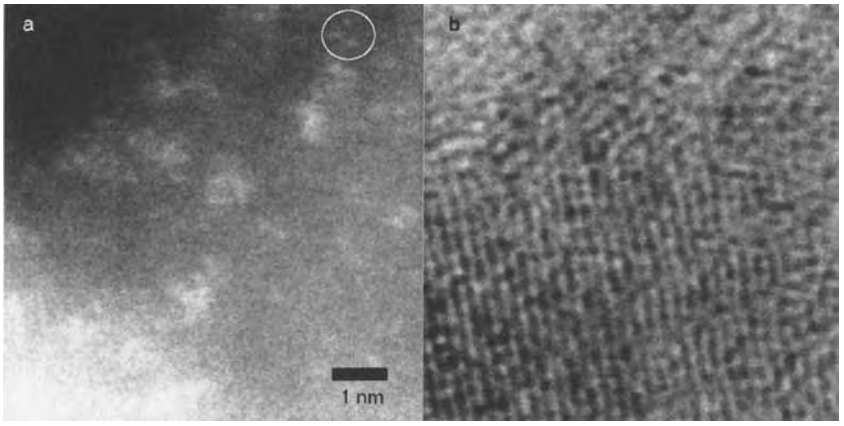


Figure 24. Images (300 kV) of a platinum catalyst on a γ-Al_2O_3 support (a) phase contrast and (b) Z-contrast, obtained simultaneously. The catalyst is seen to comprise of platinum dimers and trimers with some unresolved three-dimensional clusters.

Such images can only be obtained with the 300 kV scanning transmission electron microscope. The higher accelerating voltage reduces beam broadening by a factor of three compared to a 100 kV STEM instrument, and its probe size is almost a factor of two smaller. The probe is successfully able to image nanometer-sized clusters buried hundreds of angstroms deep inside the SiO_2. Accurate information on cluster size is essential to solve the issue of whether quantum confinement effects are playing a major role in these materials.

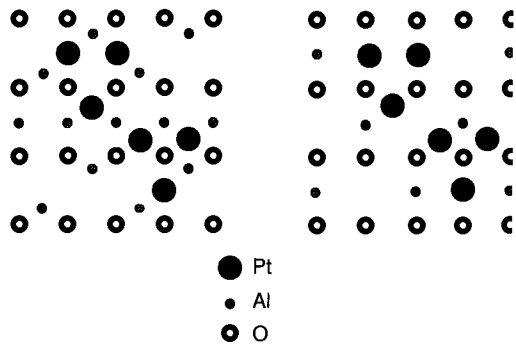


Figure 25. Possible surface configurations of platinum on the two {110} surfaces of γ-Al_2O_3, deduced from Fig. 24.

2.3.6.4 Catalysts

The ultimate example of *Z*-contrast sensitivity to date is the ability to image single platinum atoms on a γ-Al_2O_3 support, as seen in Fig. 24. The small bright dots are visible well above the noise level of the substrate, and are arranged as dimers, trimers, and other configurations. The bright field image in this case shows strong {222} fringes, and the orientation of the surface can be deduced from its Fourier transform. It is now possible to find surface sites that explain the observed spacings and orientations, as shown in Fig. 25 [58].

Such information is unobtainable by scanning tunneling microscopy since the support is insulating. Indeed, very little is known about the surface of γ-Al_2O_3 at all, but such information is essential for understanding the mechanisms of cluster formation. Again, images such as these immediately suggest avenues for theoretical studies. Figure 26 shows bright field

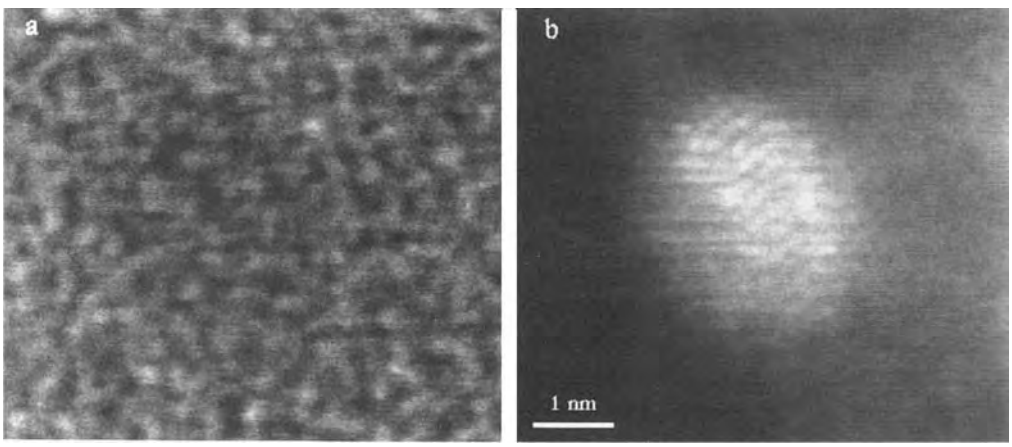


Figure 26. (a) Phase contrast and (b) *Z*-contrast images of a rhodium catalyst on γ-Al_2O_3 obtained with a 300 kV scanning transmission electron microscope.

and dark field images of a small, three-dimensional rhodium particle on γ-Al_2O_3. While the particle is barely detectable in the bright field image because of the amorphous speckle pattern from the carbon support, its internal surface is clear from the *Z*-contrast image. It is also possible to begin to deduce its external facet and step structure, again suggesting avenues for theoretical studies.

One of the most exciting possibilities would be to examine a set of samples following different chemical treatments to try to distinguish the atomic scale morphology of the catalytically active species itself. Theoretical modeling may then be able to reveal the actual reaction pathway.

Acknowledgments

The authors are grateful to A. G. Norman, V. P. Dravid, C. W. White, J. G. Zhu, D. R. Liu, R. Gronsky, and Z. Weng-Sieh for provision of samples, and to T. C. Estes and J. T. Luck for technical assistance. This research was sponsored by the Division of Materials Sciences, US Department of Energy, under contract DE-AC05-84OR21400 with Lockheed Martin Energy Systems, and in part by the Laboratory Directed Research and Development Fund of Oak Ridge National Laboratory (ORNL) and appointments to the ORNL Postdoctoral Research Program administered by the Oak Ridge Institute for Science and Education.

2.3.7 References

[1] O. Scherzer, *J. Appl. Phys.* **1949,** *20*, 20.
[2] J. W. Goodman, *Introduction to Fourier Optics*, McGraw Hill, New York **1968**.
[3] S. F. Gull, J. Skilling, *IEE Proc.* **1984,** *131F*, 646.
[4] A. V. Crewe, J. Wall, J. Langmore, *Science* **1970,** *168*, 1338.
[5] J. M. Cowley, *Ultramicroscopy* **1976,** *2*, 3.
[6] C. Colliex, C. Mory in *Quantitative Electron Microscopy* (Eds.: J. N. Chapman, A. J. Craven), Scottish Universities Summer School in Physics, Edinburgh, **1984,** p. 149.
[7] D. E. Jesson, S. J. Pennycook, *Proc. R. Soc. London A* **1993,** *441*, 261.
[8] A. V. Crewe, J. P. Langmore, M. S. Isaacson in *Physical Aspects of Electron Microscopy and Microbeam Analysis* (Eds.: B. M. Siegel, D. R. Beaman), Wiley, New York, **1975,** p. 47.
[9] M. S. Isaacson, M. Ohtsuki, M. Utlaut in *Introduction to Analytical Electron Microscopy* (Eds.: J. J. Hren, J. I. Goldstein, D. C. Joy), Plenum Press, New York, **1979,** p. 343.
[10] A. Howie, *J. Microsc.* **1979,** *117*, 11.
[11] C. R. Hall, P. B. Hirsch, *Proc. R. Soc. London A* **1965,** *286*, 158.
[12] D. M. Bird, Q. A. King, *Acta Crystallogr. Ser. A* **1990,** *46*, 202.
[13] P. A. Doyle, P. S. Turner, *Acta Crystallogr. Ser. A* **1968,** *43*, 390.
[14] D. E. Jesson, S. J. Pennycook, *Proc. R. Soc. London A* **1995,** *449*, 273.
[15] B. E. Warren, *X-Ray Diffraction*, Dover, New York, **1990**.
[16] H. Rose, *Optik* **1976,** *45*, 139, 187.
[17] R. H. Ritchie, A. Howie, *Phil. Mag. A* **1988,** *58*, 753.
[18] P. E. Batson, *Ultramicroscopy* **1992,** *47*, 133.
[19] M. J. Whelan, *J. Appl. Phys.* **1965,** *36*, 2099.
[20] V. W. Maslen, C. J. Rossouw, *Phil. Mag. A* **1983,** *47*, 119.
[21] V. W. Maslen, C. J. Rossouw, *Phil. Mag. A* **1984,** *49*, 735.
[22] C. J. Rossouw, V. W. Maslen, *Phil. Mag. A* **1984,** *49*, 749.
[23] L. J. Allen, C. J. Rossouw, *Phys. Rev. B* **1990,** *42*, 11644.
[24] L. J. Allen, *Ultramicroscopy* **1993,** *48*, 97.
[25] V. W. Maslen, *J. Phys. B* **1983,** *16*, 2065.
[26] C. J. Rossouw, V. W. Maslen, *Ultramicroscopy* **1987,** *21*, 173.
[27] H. Kohl, H. Rose, *Adv. Electron. Electron Phys.* **1985,** *65*, 2065.
[28] P. G. Self, P. Buseck, *Phil. Mag.* **1983,** *48*, L21.
[29] A. J. Bourdillon, *Phil. Mag.* **1984,** *50*, 839.
[30] S. J. Pennycook, *Ultramicroscopy* **1988,** *26*, 239.
[31] O. F. Holbrook, D. M. Bird in *Proc. Microscopy and Analysis 1995*, Jones and Begall, New York **1995,** p. 278.
[32] P. E. Batson, *Nature* **1993,** *366*, 727.
[33] N. D. Browning, M. F. Chisholm, S. J. Pennycook, *Nature* **1993,** *366*, 143.
[34] M. F. Chisholm, S. J. Pennycook, R. Jebasinski, S. Mantl, *Appl. Phys. Letts.* **1994,** *64*, 2409.
[35] S. J. Pennycook, D. E. Jesson, *Phys. Rev. Lett.* **1990,** *64*, 938.

[36] S. J. Pennycook, D. E. Jesson, *Ultramicroscopy* **1991,** *37*, 14.
[37] S. J. Pennycook, D. E. Jesson, *Acta Metall. Mater.* **1992,** *40*, S149.
[38] R. F. Loane, P. Xu, J. Silcox, *Ultramicroscopy* **1992,** *40*, 121.
[39] J. Fertig, H. Rose, *Optik* **1981,** *59*, 407.
[40] R. F. Loane, E. J. Kirkland, J. Silcox, *Acta Crystallogr. Ser. A* **1988,** *44*, 912.
[41] A. J. McGibbon, S. J. Pennycook, J. E. Angelo, *Science* **1995,** *269*, 519.
[42] A. J. McGibbon, S. J. Pennycook, *J. Microsc.* (in press).
[43] J. Hornstra, *J. Phys. Chem. Solids* **1958,** *5*, 129.
[44] M. F. Chisholm, S. J. Pennycook in *Proceedings Microscopy and Microanalysis 1996*, San Franciso Press **1996,** p. 332.
[45] O. L. Krivanek et al., *Phil. Mag.* **1977,** *36*, 931.
[46] M. F. Chisholm, M. Mostoller, T. Kaplan, S. J. Pennycook, *Phil. Mag.* (in press).
[47] A. Bourret, J. L. Rouvire, *Springer Proc. Phys.* **1989,** *35*, 8.
[48] A. P. Sutton, V. Vitek, *Phil. Trans. Roy. Soc. A* **1983,** *309*, 1.
[49] M. M. McGibbon et al., *Science* **1995,** *266*, 102.
[50] N. D. Browning et al., *Interface Science* **1995,** *2*, 397.
[51] M. M. McGibbon et al., *Phil. Mag.* (in press).
[52] M. F. Chisholm, S. J. Pennycook, *Nature* **1991,** *351*, 47.
[53] D. Dimos, P. Chaudhari, J. Mannhart, *Phys. Rev. B* **1990,** *41*, 4038.
[54] Z. G. Ivanov et al., *Appl. Phys. Lett.* **1990,** *59*, 3030.
[55] N. D. Browning, P. D. Nellist, D. P. Norton, S. J. Pennycook, *Nature* (submitted).
[56] C. C. Tsuei et al., *Science* **1996,** *271*, 329.
[57] J. G. Zhu, C. W. White, J. B. Budai, S. P. Withrow, Y. Chen, *J. Appl. Phys.* **1995,** *78*, 4386.
[58] P. D. Nellist, S. J. Pennycook, *Science* (submitted).

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

2.4 Scanning Auger Microscopy (SAM) and Imaging X-Ray Photoelectron Spectroscopy (XPS)

2.4.1 Introduction

The properties of solids are related to their electronic structure on the one hand, and to the presence of defects such as impurities, vacancies, dislocations, grain boundaries, etc. on the other hand. The surface is not thought of as containing lattice defects similar to those just mentioned; nevertheless it is also very influential on the properties of solids. Indeed, the space lattice, which is a three-dimensional infinite periodic array of points, is the geometrical representation of a perfect crystal and departures from this mathematical model are by definition lattice defects. The discontinuity which constitutes the surface is therefore a lattice defect. As a result of the presence of unsaturated bonds, the surface induces properties in the solid which are the direct result of its existence. In this manner the surface is very similar to the other lattice defects already mentioned above and abundantly discussed in this Handbook.

The surface is also the agent through which the solid interacts with its surroundings and is therefore of great importance in, for example, microelectronics, catalysis, corrosion, etc. It has therefore been intensively studied since the early 1960s when UHV (ultrahigh vacuum) became readily available. Indeed, to obtain clean surfaces, vacua in the 10^{-10} mbar range are necessary and these can now routinely be obtained.

When studying a surface one is interested in its structure and chemical composition, that is, the chemical bonding at the surface and in the surface layer. Yet, when studying a surface one does not restrict oneself to the uppermost layer only. It is important to examine further layers, even to a depth of a few nanometers thick, because properties related to the surface are also influenced by the subsurface layers.

A large number of techniques for studying these properties have been developed and, in this chapter, two of the oldest, but most widely used, techniques are described: namely, Auger electron spectroscopy (AES) and X-ray photoelectron spectroscopy (XPS) or ESCA. Both techniques are used to study the chemical composition of the surface and their information depth is limited to a few atomic

layers, making them ideally suited for this purpose. AES was developed about 1967 first using the LEED retarding grids as electron spectrometer, although soon afterwards the cylindrical mirror analyzer (CMA) was introduced and became commercially available. AES has now evolved into the most widely used surface analysis technique. It is found not only at universities and research institutes, but also in industry. XPS became commercially available around 1970, following the pioneering work of Siegbahn's group at Uppsala [5]. Its potential was immediately recognized and, consequently, XPS also evolved into a widely used surface analysis technique, especially after it was realized that the information depth of both techniques is practically equal. For a long time, however, the limited lateral resolution of XPS constituted a drawback for its use in, for instance, microelectronics. The recent developments described in this chapter have, however, drastically altered this perspective.

Lateral resolution in AES was less of a problem than in XPS because, in 1972, it was suggested that SEM and AES should be combined into a single instrument. This resulted in AES spectra with high lateral resolution, as will be further described. The scanning Auger microscope (SAM) consequently became available by the late 1970s and has subsequently found extensive use, particularly in industrial laboratories. Imaging XPS was introduced in the mid-1980s, and benefited from developments in the field of X-ray focalization and monochromatization and improved electron detection capabilities such as multichannel plates.

In broad terms, Sec. 2.4.2 of this Chapter summarizes the general principles and techniques of AES and XPS. Section 2.4.3 describes, in some detail, the experimental aspects of both scanning Auger microscopy and imaging XPS. Finally, Sec. 2.4.4 describes the characteristics of the images obtained by these techniques.

2.4.2 Basic Principles of Auger Electron Spectroscopy (AES) and X-Ray Photoelectron Spectroscopy (XPS)

2.4.2.1 Auger Electron Spectroscopy (AES)

Many review papers in the literature describe in detail the different aspects of the Auger process. Consider, for example, the review paper by Fiermans and Vennik [1]. Only the main aspects necessary to understand what follows are summarized here.

The Auger Process

Auger electrons emerge from a solid when the latter is excited with energetic electrons (in the keV range), X-rays, ions, protons, etc. They are a consequence of radiationless rearrangements of the electrons in atoms in which a core hole has been created by the exciting radiation. The energy released during this radiationless rearrangement is transferred to an electron which emerges from the solid with a certain kinetic energy, the atom being left behind in a doubly ionized state. This electron is called the Auger electron, after the French

physicist, Pierre Auger, who first described this process in the 1920s.

Auger Notations

Taking into account the spin–orbit splitting, the electrons are arranged in an atom in the shells K(1s), $L_1(2s)$, $L_2(2p_{1/2})$, $L_3(2p_{3/2})$, $M_1(3s)$, $M_2(3p_{1/2})$, $M_3(3p_{3/2})$, ..., a notation used in X-ray spectroscopy. Considering, for instance, the removal of a 1s electron from the K-shell, one has radiative KL_2 (and KL_3) transitions giving rise to X-ray photons, and a number of radiationless transitions where the energy is transferred to a second electron. This can also be an electron in the L_2 (or L_3) shell and the Auger electron is then called the KL_2L_2-electron. This, however, is not the only possibility and a whole series of KLL-processes has to be considered, each giving rise in L–S coupling to a number of spectroscopic terms ($^1S_0$; $^1P_1$, $^3P_{2,1,0}$; $^1S_0$, $^3P_{2,1,0}$, $^1D_2$ for the $2s^02p^6$, $2s^12p^5$ and $2s^22p^4$ final states, respectively). Some of these terms are parity forbidden ($2s^22p^4$: ^3P for instance) and in j–j coupling only six terms remain, namely: KL_1L_1, KL_1L_2, KL_1L_3, KL_2L_2, KL_2L_3 and KL_3L_3.

In practice one often finds a set of terms determined by an intermediate coupling scheme. The notation including the symbol of the spectroscopic term considered, is generally adopted, e.g. $KL_{2,3}L_{2,3}(^1D)$. The same notation is used for Auger processes with initial holes in L, M, N, ... shells (e.g., $L_3M_{4,5}M_{4,5}(^1D)$, etc.).

When one or more levels participating in the process are situated in a composite valence band, the symbol V is used (e.g., KVV, L_3VV, $L_3M_{2,3}V$, etc.).

Coster–Kronig transitions are Auger processes involving sublevels of the same shell (e.g., $L_1L_{2,3}M_{2,3}$). Their rate is strongly dependent on the energy separation between the sublevels considered. Super Coster–Kronig transitions are processes occurring in the same shell (e.g., $M_1M_2M_{4,5}$, etc.).

Energies of the Auger Transitions

When electron holes (e.g., in a photoemission or in an Auger experiment) are created, a charge rearrangement takes place in the atom. It is, therefore, necessary to include in the Auger kinetic energy formula for an Auger transition involving, for instance, the j, k and l levels, terms related to multiplet splitting and electronic relaxation (see [2] for a more detailed account):

$$E_{\text{kin}}(jkl;X) = E_b(j) - E_b(k) - E_b(l) - F(kl;X) + R_c \qquad (1)$$

wherein X is the multiplet state resulting from the coupling of the two holes k and l in the final state. The E_b-terms are ground-state one-electron binding energies, for which experimental values can be used. In this way initial state chemical effects and one-hole electronic relaxation phenomena are automatically taken into account.

The two-electron F-term accounts for the fact that within the two-step model considered for the Auger process, upon emission of the Auger electron l, already one electron, the electron k, is missing. As a result, the binding energy of the electron l increases with respect to the ground-state value, $E_b(l)$. The R_c-term is a supplementary term, called cross relaxation energy [2], which contains both an atomic and an extra-atomic contribution. It primarily accounts for the more repulsive medium

that is created for the electron l due to the relaxation that has taken place towards the electron hole k.

In this formula the F and R_c terms have to be calculated, which sometimes can be done for atomic and quasiatomic Auger transitions. Experimentally, however, a semiempirical formula is preferred, using experimentally determined and published electron binding energies.

Consider, for instance, an Auger transition KL_2L_3. To a first approximation, the following expression for kinetic energy is readily obtained:

$$E_{kin} = E_K - E_{L_2} - E_{L_3(L_2)} \qquad (2)$$

Self-consistency is introduced in such semiempirical expressions through the Chung and Jenkins formula [3]:

$$E_{kin} = E_K(Z) - \tfrac{1}{2}[E_{L_2}(Z) + E_{L_2}(Z+1)] - \tfrac{1}{2}[E_{L_3}(Z) + E_{L_3}(Z+1)] \qquad (3)$$

Using Eq. (3), kinetic energy values have been tabulated by Coghlan and Clausing [4] and these are now widely used.

The Auger energies measured are kinetic energies referred to the vacuum level of a spectrometer (Fig. 1). Since binding energies generally are referred to the Fermi level, it is necessary to add the spectrometer work function Φ_{anal} to the measured Auger energy to obtain values comparable with those calculated.

The spectrometric work function is an unknown factor and, furthermore, can change from one experiment to another. Normally, the energy scale is calibrated using suitably chosen primary peaks and once the system is calibrated it is tacitly assumed that it is not modified in the course of experiments.

Since the primary energy of the exciting electrons in AES generally is in the range

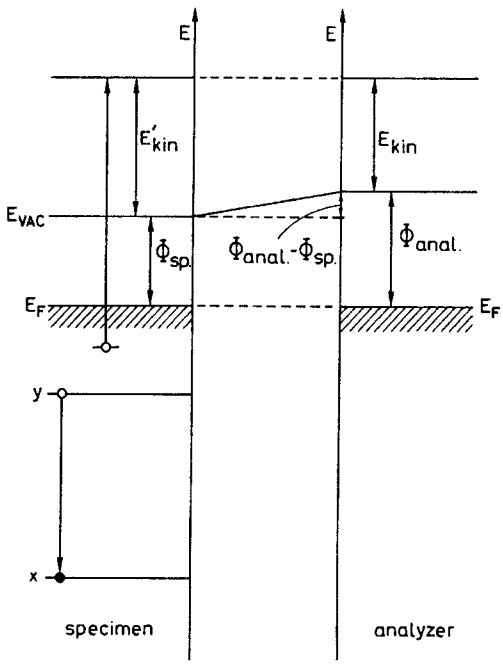


Figure 1. Energy-level diagram showing the relationship between actual and measured kinetic energies.

3–10 keV, only K-shell core holes of the light elements (C, O, ...) can be created, giving rise to the K-series Auger lines. For medium and heavy elements one excites the L-, M-, ... series Auger lines. Summarizing, the observed Auger electron kinetic energies are generally situated in the energy range 0–2000 eV, limiting the information depth. Indeed, the escape depth of these electrons, without energy loss, is limited to a few atomic layers. This point will be discussed in detail in the paragraph on quantitative analysis.

Chemical Shift and Line Shape of Auger Lines

Depending on the chemical environment of the ion in which the transition occurs, shifts in the positions of Auger peaks are commonly observed. The so-called

chemical shifts are generally larger in AES than in XPS (see below). In AES the observed shifts are a sum of binding energy shifts (the true chemical shift) and the change in total relaxation energy. Unfortunately, the latter is difficult to determine for most compounds. These chemical shifts are important when imaging the chemical composition. Therefore, this detail is discussed in the following account. The shape of an Auger line is also strongly dependent on the chemical environment and can also be used for chemical composition imaging as will be shown.

Experimental Aspects of AES

(a) *Energy analyzers*: These are abundantly discussed in the second paragraph of this chapter.

(b) *Sample preparation*: A specimen prepared in normal atmospheric conditions is covered by a layer of contamination. This layer gives rise to relatively strong carbon and oxygen spectral lines and reduces the intensity of the Auger lines of the elements present in the solid. To increase the signal-to-background ratio, the surface is normally cleaned by Ar^+-bombardment. This, however, carries in it the danger of preferential sputtering of the solid under study, with the inherent danger of altering relative intensities in an elemental quantitative analysis. A better way to obtain atomically clean surfaces is via UHV in situ preparation involving scraping, fracture, cleavage, etc. However, this is not always possible.

Beam effects: Certain compounds are subject to rather severe beam effects in AES. The primary electron beam can, for example, cause surface decomposition, electron assisted adsorption, electron assisted desorption, or induce formation of compounds (e.g., carbide formation on metals), which have all to be taken into account when performing the surface chemical composition analysis. The Ar^+ ion beam bombardment mentioned above is commonly used to record depth profiles. In such an analysis the specimen is subjected to both an energetic ion and electron beam bombardment, generally resulting in a modification of the original surface composition. This has clearly to be borne in mind when interpreting these profiles.

Another beam effect concerns nonconducting specimens. To obtain reliable Auger spectra, the specimens should be conducting and at ground potential. If this is not the case charging will immediately occur, rendering impossible any further measurement.

2.4.2.2 X-Ray Photoelectron Spectroscopy (XPS)

The XPS Process

The XPS process is nothing more than the photoelectric process, whereby an electron is emitted from a solid when the latter is irradiated with X-rays. In the emitting atom an electron hole is created and charge rearrangement will occur because a nonrelaxed one-electron hole system is not an eigenstate of the corresponding Hamiltonian. A charge rearrangement will take place to screen-off the positive hole, thereby lowering the total energy of the system. As the charge rearrangement generally occurs in a time interval that is

comparable with the time of the photoemission process itself, the reduction of the total energy, with respect to the frozen orbital situation, leads to an increase of the kinetic energy of the outgoing photoelectron. Since the binding energy of an electron is defined as the difference between the kinetic energy of the incoming photon and the kinetic energy of the outgoing photoelectron, that is,

$$E_{\text{kin}} \text{ (outgoing } e^-\text{)} = h\nu \text{ (incoming photon)} - E_b(i) \text{ (binding energy of the } i\text{th electron)}$$

the increase of the kinetic energy of the detected electron results in a reduction, of the order of a few eV, of the effectively experimentally observed binding energy. These electronic screening effects occur in free atoms as well as in the condensed phase. As a result of the so-called extra-atomic relaxation, screening will normally be larger for the condensed phase than for free atoms. Consequently, in the discussion of chemical shifts (below), it should be borne in mind that these are not only determined by initial state charge transfer phenomena, but also by final state electronic screening effects.

When measuring the kinetic energy of the photoelectron with an analyzer, the work function, Φ, of the latter must be taken into account, binding energies being referred towards the Fermi-level: $E_{\text{kin}} = h\nu - E_b - \Phi$.

For photoelectrons, a notation based on an X-ray notation, using the spectroscopic level (e.g., C1s, $Si2p_{1/2}$, $Si2p_{3/2}$, etc.) is adopted. To a first approximation, the initial states before photoemission are characterized by completely filled shells, while after photoemission one electron hole is present. One consequently only expects the corresponding spin–orbit splitting spectroscopic terms, explaining the notation used above. If, however, the electron hole thus created can couple with other holes already present, multiplet splitting occurs and the photoline becomes much more complex than the single lines just described.

Since the FWHM value of the photoline is determined by, among other factors, the width of the exciting X-ray photon, narrow monoenergetic soft X-ray lines such as AlKα (1486.6 eV) or MgKα (1253.6 keV) are used as primary excitation. The kinetic energies of the photolines are consequently limited to these values. The escape depth of the photoelectrons is therefore comparable to the one discussed for Auger electrons. This point will be dealt with in more detail in Sec. 2.4.2.4 of this Chapter.

In an XPS spectrum, one notices the presence of core level photolines next to the valence band structure. In such a spectrum one normally can also clearly distinguish the presence of Auger lines. Indeed, upon creation of the core holes, deexcitation readily occurs, giving rise to the emission of Auger electrons, which are measured in the spectrometer (Fig. 2).

Determining binding energies and consequently identifying the chemical composition of a surface layer is the main subject of XPS. It is also known under the popular acronym ESCA (electron spectroscopy for chemical analysis) [5].

Analysis is based on the fact that each element has a unique spectrum and can

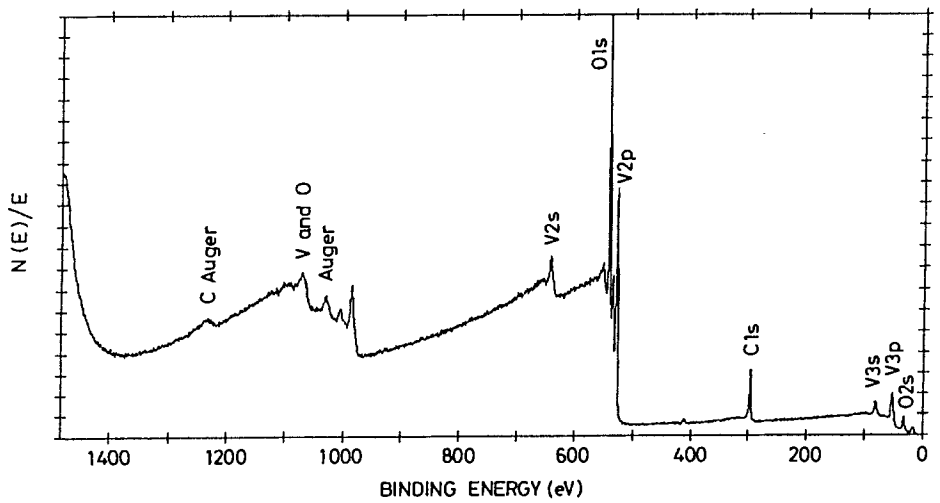


Figure 2. XPS spectrum of V_2O_5 obtained with a CHA/lens combination (CHA in retarding mode).

therefore be used as a kind of fingerprint. It is also straightforward that a mixture of elements gives rise to a mixture of these unique spectra and, to a first approximation, quantitative data can be obtained from the peak areas. In general, the identification of the peaks poses no severe problems, providing that one is aware of the fact that satellites such as plasmon losses, shake-up, and shake-off satellites may be present.

Beam Effects

Beam effects are much less severe in XPS than in AES. Indeed, the soft X-ray beam induces only minor photoreduction effects on certain compounds. Charging is also less of a problem when measuring non-conducting samples. In XPS the sample charges positively as a result of the photoemission, inducing a shift towards higher binding energies and a general broadening of the photolines. This effect can be compensated by flooding the sample with very low energy electrons produced by an auxiliary electron flood gun.

2.4.2.3 Quantitative Analysis in AES and XPS

There are two possible approaches to the problem of quantitative analysis: (i) a purely formal ab initio approach, or (ii) a semi-empirical approach. The latter is routinely installed in most AES and XPS instruments. Inevitable approximations in these routines can lead to considerable errors. It should also be pointed out that AES and XPS are not trace analysis techniques, their detection limits being in the range 0.1–1% for most elements. The semi-empirical approach was discussed in detail a few years ago by Nebesny et al. [6] and we will rely heavily on this excellent paper in our discussion.

An electron beam excited Auger signal is given by a formula, introduced by Seah

[7] (and using the notation of Nebesny et al. [6]):

$$I_A = I_0 \times \sigma_{A(E_p)} \times (1 + r_{M(E_A,\alpha)}) \times T_{(E_A)} D_{(E_A)} \times \int_0^\infty N_A(z) \times \exp[-z/\lambda_{M,E} \cos\theta]\, dz \quad (4)$$

where I_0 is the primary electron beam current ($A\,cm^{-2}$), $\sigma_{A(E_p)}$ is the electron impact ionization cross section at primary energy E_p, the factor $(1 + r_{M(E_A,\alpha)})$ refers to Auger electrons at energy E_A produced by backscattered electrons in which the term $r_{M(E_A,\alpha)}$ is the backscattering coefficient at angle α to surface normal, $T_{(E_A)}$ is the transmission efficiency of analyzer at E_A (Auger kinetic energy), and $D_{(E_A)}$ is the detector efficiency. The integral contains the analyte concentration (N_A), modified by an exponential decay term, with decay constant λ, the electron escape depth at energy E_A, and θ is the angle of analysis to surface normal.

The escape depth is defined as the depth from which an Auger electron (or for that matter a photoelectron) can escape from the solid without undergoing inelastic collisions whereby it may lose part of its energy and disappear in the background. Experimentally determined values of λ give rise to the so-called 'universal escape depth' curve from which it follows that the escape depth for electrons with kinetic energy limited to 2000 eV, is only a few atomic layers (0–2 nm) (Fig. 3). Assuming

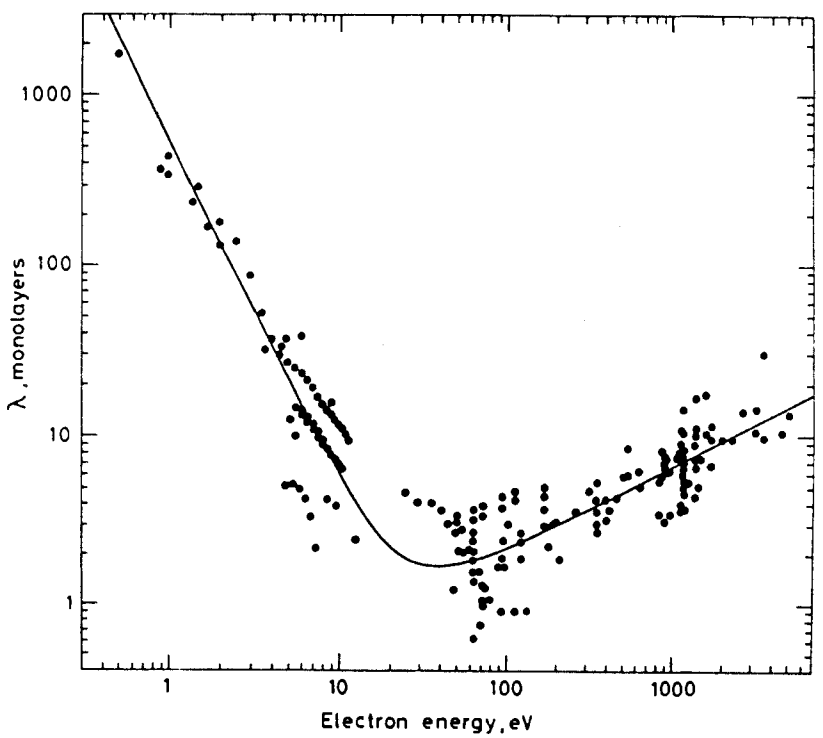


Figure 3. Escape depth in function of electron energy. Adapted from [7]. Reproduced by permission of M. P. Seah and John Wiley & Sons Ltd.

Auger emission occurs within $z = 5\lambda$, an information depth limited to maximum 100 Å is obtained. This equation is an ab initio formula wherein the principal uncertainties arise from $\sigma_{A(E_p)}$, r_M, λ_{M,E_A} and the instrumentally dictated parameters $T_{(E_A)}$ and $D_{(E_A)}$.

The easiest application of this formula is for submonolayer adsorbates, where the complete formula is employed [8]. When the solid is homogeneous within the analysis depth for all the elements of interest (A and B for instance), the uncertainties can be partially canceled by considering relative atomic values. In this case one obtains:

$$\frac{I_A}{I_B} = \frac{\sigma_A(E_A)}{\sigma_B(E_B)} \times \frac{(1 + r_{M(E_A)})}{(1 + r_{M(E_A)})} \times \frac{\lambda_{M,E_B}}{\lambda_{M,E_B}} \times \frac{T_{E_A}}{T_{E_B}} \times \frac{D_{E_A}}{D_{E_B}} \times \frac{N_A}{N_B} \tag{5}$$

If T and D are known to within a certain tolerance, these terms can then be canceled. If pure element standards are available for A and B, an atomic percentage of A and B can be computed from I_A/I_B and I_A^∞/I_B^∞ (the Auger current ratio from the pure element standards) and a correction for matrix effects, a 'matrix factor' $F_{A,B}^A$. An alternative is to use standards that are near the suspected composition and have the same matrix as the unknown.

The relative atomic ratio $(N_A/N_B)_{unk}$ of the unknown is computed from the formula:

$$\left(\frac{N_A}{N_B}\right)_{unk} = \left(\frac{I_A}{I_B}\right)_{unk} \times \left(\frac{I_B}{I_A}\right)_{std} \times \left(\frac{N_B}{N_A}\right)_{std} \tag{6}$$

One obtains (see Nebesny et al. [6] for more details) for the atomic percentage X_A:

$$X_{A(unk)} = \frac{I_{A(unk)}}{\sum_i I_{i(unk)}} \times \frac{\sum_i I_{i(std)}}{I_{A(std)}} \times \frac{1}{X_{A(std)}} \tag{7}$$

where $(I_A/I_B)_{unk}$ and $(I_A/I_B)_{std}$ are Auger signal ratios for the unknown and the standard, and $(N_A/N_B)_{std}$ is the stoichiometry of the standard. Another option consists of using sensitivity factors for each element obtained from different sources. This option, however, has to be used with care since the factors can differ from one instrument to another.

As an example, the procedure employed by Physical Electronics Inc., and routinely installed in the instrument [9], can be mentioned. With sensitivity factors using, for instance, Ag as a standard, the procedure leads to a much simpler formula:

$$C_X = \frac{I_X}{I_{Ag} S_X D_X} \tag{8}$$

where C_X is the atomic concentration of element X, I_X is normally the peak-to-peak amplitude of the element X, I_{Ag} is the peak-to-peak amplitude for the Ag standard, D_X is a relative scale factor which can be calculated and S_X is the relative sensitivity of element X towards silver.

By using peak-to-peak amplitudes (discussed below) variations in the Auger line shape are neglected. For instance, when recording a depth profile, elemental concentration data using this formula can be transformed by statistical methods (factor analysis and MLCFA correlations) into more reliable chemical information [10]. In MLCFA (maximum likelihood

common factor analysis), the Auger peak shape is taken into account and the number of components building this shape is determined. Each component corresponds to a particular chemical environment of the element considered. We refer the reader to the related literature [10] for a more detailed account of this powerful approach to quantitative chemical composition determination.

For XPS similar formalisms have been developed, taking into account peak shapes and the presence of satellites. In XPS the area under the photoline is normally used for quantitative analysis; however, the question is how to obtain accurate peak areas both in AES and XPS. A particular problem concerns accurate background subtraction.

In AES the background arises from different processes [6]:

- energy loss by electrons from the source originally at higher kinetic energies;
- electrons from Auger emission at higher kinetic energy;
- intrinsic losses, mainly in free electron metals, due to relaxation effects;
- extrinsic energy losses such as plasmons.

The first two processes lead to an exponentially growing background towards lower kinetic energies, the so-called 'secondary cascade' on which the Auger electrons are superimposed.

In XPS the principal processes contributing to the background are:

- photoemission events of higher kinetic energy, leading to the 'secondary cascade';
- similar extrinsic losses due to plasmons;
- shake-up and shake-off satellites.

In both AES and XPS instruments, the background is normally subtracted following different steps: (i) dividing the spectral intensities by the kinetic energy; (ii) linearization of the log–log form of the data and subtraction; and (iii) the integral background method [11] is available in virtually all commercial instruments. In this technique the integral is calculated over the whole peak area and a calibrated curve thus obtained is subtracted from the peak area. Further details of (i)–(iii) are provided in [6]. Ideally, one could then deconvolute the experimental peak shape with an instrumental curve obtained with a primary peak at the peak value under study. However, this is not commercially available in the present instruments.

Concentrations and concentration profiles are then calculated from the peak areas obtained after carrying out the above background subtraction routine. One might, in principle, decompose the photoline into its different components when different valence states (chemical shifts) are present, and then calculate relative concentrations of the different chemical states of the element present. This approach is not routinely available in most commercial instruments as it involves rather long data acquisition times.

2.4.3 Scanning Auger Microscopy (SAM) and Imaging XPS

2.4.3.1 Basic Principles of Imaging

In microscopic imaging techniques, just as in classical optical systems, the object or

sample to be viewed is illuminated by a source, broad enough to illuminate the whole sample at once. The reflected light or emitted beam from the sample is then passed through a magnifying imaging device (cf. an optical lens) and hence a magnified image of an object or sample is obtained in which all details are seen simultaneously. In pure AES or XPS spectroscopy, the primary aim is to energy analyze either Auger and/or photoelectrons originating from a sample under the influence of the illumination or excitation of an electron or X-ray beam. It is, of course, the energy analysis that is essential to both AES and XPS, which gives the required elemental and binding energy oriented information. However, as in most energy analyzers, the energy dispersion results in a spatial displacement of the image as a function of energy (Fig. 4). The combination of energy analysis with imaging is not straightforward and in the simple scheme described one ends up with a 'picture' at the detector plane which is the convolution of energy information and image information. Such a picture is neither interpretable nor usable for simple microscopic purposes.

The simplest solution to this problem is to step down from the concept of global or simultaneous imaging of the sample and to divide it into as many pixels as needed to obtain a good quality image: in other words, to obtain a good lateral resolution. Each pixel then acts as a point source for the energy analyzer. By performing a sequential energy analysis of each pixel, a picture of the sample can be reconstructed at a fixed electron energy. The question now is how to divide the sample into pixels. If the illuminating beam can be focused and steered or scanned over the sample as with charged particle beams, a sample is easily 'pixeled'. The area of the pixel or point source, seen by the energy analyzer, is then defined by the focusing properties of the illuminating or primary beam. The location of the pixel on the sample depends on the spatial position of the primary probe and is defined by deflecting voltages on scan plates between

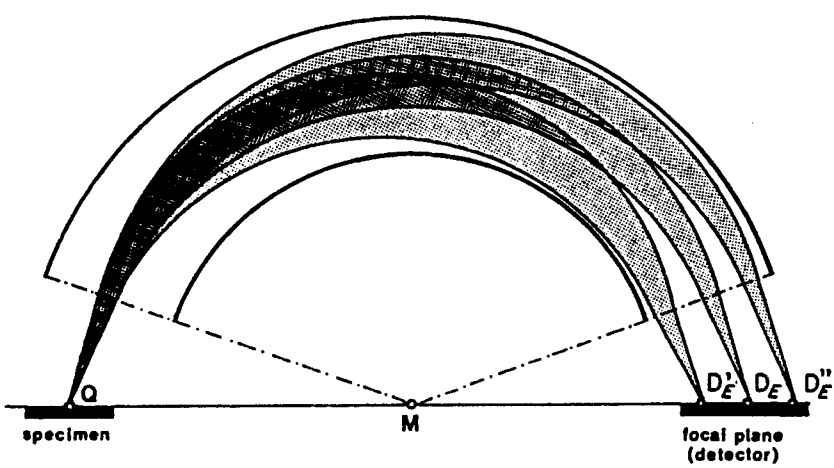


Figure 4. Energy selection gives rise to a spatial displacement of the image. Adapted from N. Gurker, M. F. Ebel and H. Ebel, *Surf. Interface Anal.* **1983,** *5*, 13 (Fig. 5). Reproduced by permission of N. Gurker and John Wiley & Sons Ltd.

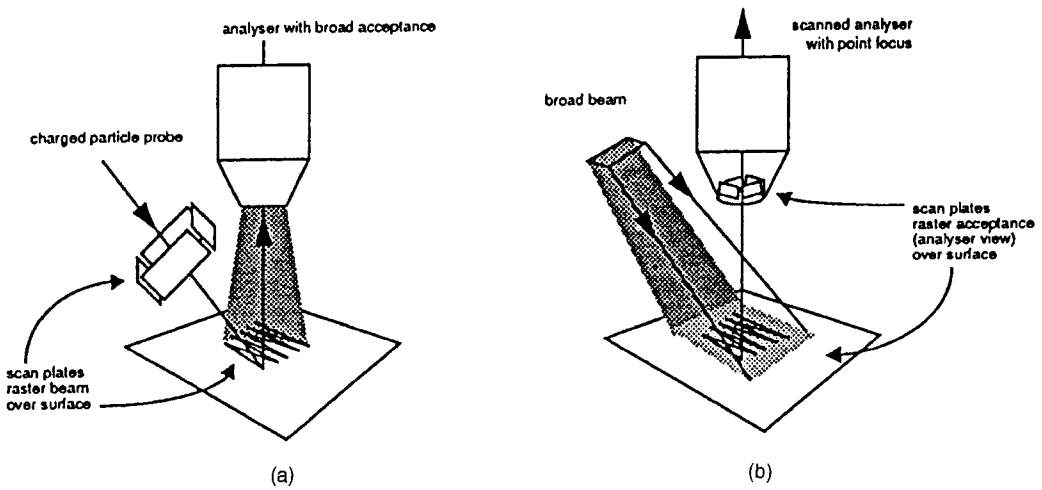


Figure 5. (a) Source defined imaging: the primary excitation source is scanned over the sample to be viewed. The analyzer, with an acceptance broad enough to see the scanned area, accepts the information of each pixel. (b) Detection defined imaging: the sample area under investigation is illuminated with the primary excitation source. The analyzer with restricted acceptance sees each pixel. Adapted from Kratos System Overview Axis 165. Reproduced by permission of Kratos Analytical Ltd.

source and sample. For this approach, an analyzer with an acceptance broad enough to see the scanned area on the sample is needed. This solution to energy dispersive imaging is called (Fig. 5a) 'source defined' imaging. If the primary beam is not steerable and hardly focusable, as is traditionally the case with, for instance, X-rays, another solution is required if charged particles are to be detected. By reducing the acceptance of the energy analyzer, the lateral dimensions of the area under investigation are reduced until the energy dispersive system sees a pixel or point source. By rastering this accepted area over the sample, using deflection plates in front of the analyzer (Fig. 5b), electrons are again sequentially collected from the different pixels. This approach to imaging is called 'detection defined' imaging and requires a broad and homogeneous primary excitation beam. As detailed later, it is between these two apparent extreme solutions that other approaches to imaging in AES and XPS have developed into powerful analytical tools.

2.4.3.2 General Aspects of Analyzers

Basically, there are three ways in which the energy of charged particles may be measured.

Retarding Field Analyzers

The idea behind the use of retarding fields is that only those particles with sufficient energy can reach a collector. Clearly, such retarding fields act as high pass energy filters and the collected particle current is given by

$$I(E_0) = K \cdot \int_{E_0}^{\infty} N(E)\,\mathrm{d}E \qquad (9)$$

where K is the instrumental constant, E_0 is the cut off energy corresponding with a retarding potential V_0, and $N(E)$ is the energy distribution of incident particles.

By modulating the retarding potential V_0 with an alternating signal ΔV_0 and by synchronously detecting the AC part of the collected current, one effectively measures the first derivative of I with respect to E_0 [12], that is,

$$\frac{dI(E_0)}{dE_0} \sim N(E_0) \tag{10}$$

In fact, this was how, in 1967, AES was introduced by using LEED optics as a retarding field analyzer. Considering the secondary electron emission distribution of Fig. 6, it should be noted that Auger peaks are rather small features. This is the reason why normally $dN(E_0)/dE$ versus E_0 spectra were recorded and amplified, using well established lock-in amplifier techniques. Synchronous detection of the second harmonics content of the current $I(E_0)$ gives a signal proportional with the required first derivative $dN(E_0)/dE_0$. In this second harmonics mode of the lock-in modulation technique, the background is greatly reduced and electronic amplification can be considerably increased. Another approach to obtaining a bandpass instead of a high pass characteristic is to reaccelerate and selectively collect those particles which pass through the decelerating field at minimum kinetic energy. In other words, after passing the decelerating high pass filter, particles must subsequently pass through a low pass energy filter.

Spectrometers of this type [13–16] are available commercially for both AES and

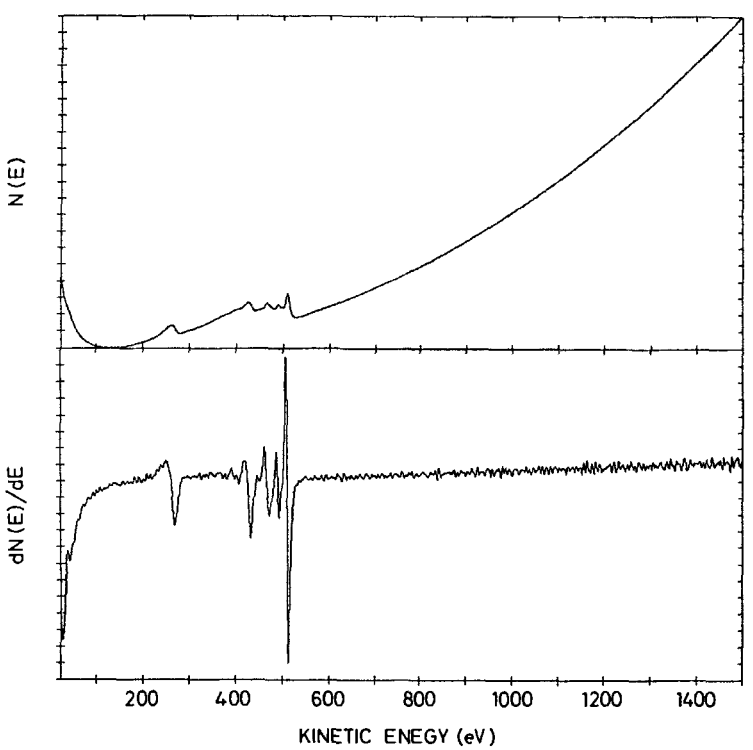


Figure 6. Auger spectrum of V_2O_5.

XPS and are characterized by a large solid acceptance angle Ω and by a large étendue λ. The étendue λ, as proposed by Heddle [17], is defined as the product of accepted area S and of accepted solid angle Ω

$$\lambda = S\Omega \tag{11}$$

Due to the lack of good lateral resolution, these spectrometers can only be used for imaging purposes in the source defined mode, that is, with a well-defined primary excitation beam scanning the sample surface and yielding the required lateral resolution.

Deflection Type Analyzers

A second approach is by deflection in electric and/or magnetic fields. Historically, the oldest application of energy analyzers can be traced back to β-spectroscopy, in which different magnetic deflection schemes were developed. However, during recent decades, the pure electrostatic energy analyzer has become the more popular instrument for several reasons. Practical reasons, for instance, are to be found in the less bulky construction as compared with most magnetic instruments and also in the fact that pure electrostatic devices are better adapted for UHV environments as used in AES and XPS experiments.

There are also more fundamental reasons for the popularity of electrostatic analyzers. For example, it is much easier to define the boundary conditions in electrostatic systems as compared with magnetic layouts. Shielding magnetic fields is much more of a problem than shielding electrostatic fields. In pure electrostatic deflection devices, the trajectories of a particle are independent of mass and charge [18]. Mass and charge only intervene in the transit time of the particle along the followed ray path. Clearly, this property is of value in the energy analysis of ions having different masses and (multiple) charges; only the ratio of electrode potentials to the accelerating potential of the particles remains constant [18].

A strong impetus to the development of electrostatic energy analyzers has originated from ion beam research. Many different types of deflecting electrostatic energy analyzer have been developed and applied successfully. Without elaboration, the most important types are as follows:

- Parallel plate analyzers [19] and fountain analyzers [20]. Both are characterized by a uniform electric field between parallel electrodes.
- Cylindrical sector fields. These are best known through a special focusing configuration, namely, the 127° sector analyzer [21, 22], and cylindrical mirror analyzers (CMA) [23–29]. Both types are characterized by an inverse first power of r force field. The cylindrical mirror analyzer was first used in AES by Palmberg et al. [30] in 1969 and then became a standard tool in AES. It also has good transmission (i.e., ratio of emergent particle flux to incoming particle flux), of the order of 10% for moderate resolution (0.3%).
- Hemispherical analyzers [17, 31, 32]. These are characterized by an inverse second power of r force field.

Time of Flight (TOF) Analyzers

In contrast with ion spectroscopic techniques, time of flight (TOF) methods are not really used for energy analysis of

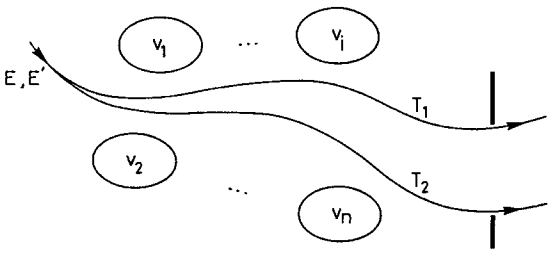


Figure 7. Trajectories T_1 and T_2 are the limiting trajectories through the output aperture.

electrons in AES or XPS. Due to the much smaller mass of an electron as compared with ions and taking into account the required relative energy resolution (~0.1%), timing electronics are not readily available for a time of flight setup with realistic dimensions of the order of 1 m maximum. Furthermore, TOF techniques require fast pulsed excitation sources and, in any case, the required electronics are not readily available.

2.4.3.3 Energy Resolution of Deflecting Electrostatic Analyzers

Of course, an important figure of merit of an energy analyzer is the energy resolution although for deflecting electrostatic analyzers one is always confronted with the relative resolution,

$$R = \frac{\Delta E}{E} \tag{12}$$

The reason for this can be briefly explained as follows. Consider a system with n completely arbitrary shaped electrodes carrying potentials $V_1, \ldots, V_n$. This potential V_i of the i-th electrode is measured with respect to a reference potential V_0 and the measured potential differences are given by

$$v_i = V_i - V_0 \tag{13}$$

The reference potential is chosen such that at V_0 the particles injected into the system under consideration are at rest, that is, at zero kinetic energy. This means that the kinetic energy is given by

$$E = q(V - V_0) = qv \tag{14}$$

if the particle is accelerated by the potential V. As already mentioned [18], the trajectories only depend on the ratio of electrode potentials to the accelerating potential. This means that in Fig. 7 trajectory T_1 is characterized by

$$\frac{v}{v_i} = K_i \quad 1 \leqslant i \leqslant n \tag{15}$$

with v the reduced potential given by Eq. (14) and corresponding with a kinetic energy E of a particle entering the electrode system. Similarly, a particle with a kinetic energy E'

$$E' = q(V' - V_0) \tag{16}$$

will follow a trajectory T_2 through the electrode system and this trajectory is characterized by

$$\frac{v'}{v_i} = K_i' \quad 1 \leqslant i \leqslant n \tag{17}$$

From Eqs. (14) and (16) it follows that

$$\frac{v'}{v} = \frac{E'}{E} = K_0 \tag{18}$$

If $\Delta E = E' - E$ it is obvious that $R = \Delta E/E$ is also constant.

In other words, for given electrode potentials the trajectories T_1 and T_2 depend on the respective energies E and E'. If those trajectories are the limiting trajectories through an exit aperture as is the case for an energy spectrometer, the relative energy resolution $R = \Delta E/E$ is a constant and depends on the system geometry.

Clearly, the higher the transmitted energy E, the higher the absolute energy resolution or energy spread ΔE. An important consequence of the constant relative resolution, $\Delta E/E$, is that when measuring an energy distribution $N(E)$ at the output of the analyzer it is not simply a measure of current, that is,

$$I(E) \sim N(E) \tag{19}$$

but, instead

$$I(E) \sim \Delta E \times N(E) \tag{20}$$

or

$$I(E) \sim E \times N(E) \tag{21}$$

To eliminate this problem, in other words to measure an energy distribution with constant ΔE, it is necessary to shift the energy E of the incoming electrons, either by retarding or accelerating them in an appropriate electric field, by an amount E' such that the final energy is always at the fixed pass energy E_0 of the analyzer, that is,

$$E - E' = E_0 \tag{22}$$

As the pass energy of the analyzer is fixed so is the absolute energy spread ΔE during the measurement. As far as the electrostatic energy dispersive system is concerned, the measured current at the output is simply proportional to $N(E)$; however, the retardation also influences the output current of the complete spectrometer. Again, the background can be eliminated by measuring the first derivative of $N(E)$ with respect to E. By modulating the deflecting voltages of the analyzer, this can be achieved by synchronously demodulating the output signal and using the first harmonics content. As in the nonretarding mode, the output current is, according to Eq. (21), proportional to the product $E \times N(E)$, the first derivative now becomes

$$\frac{dI}{dE} \propto E\frac{dN}{dE} + N(E) \tag{23}$$

To a good approximation, this reduces to

$$\frac{dI}{dE} \propto E\frac{dN}{dE} \tag{24}$$

which again is obtained by measuring the first harmonic with lock-in detection techniques. The sensitivity of these detecting techniques was greatly enhanced by using electron multipliers. In these analogue detection layouts, use is normally made of high current type electron multipliers. In contemporary instrumentation, analogue detection is frequently supplanted by fast pulse counting techniques in which each individual electron hitting the electron multiplier is detected and counted. This results in much better signal-to-noise ratios as compared with the described analogue techniques. As will be clear, pulse counting yields directly either $N(E)$ or $E \times N(E)$ as a function of E, depending on whether the retarding or nonretarding mode was used. Background subtraction or other data treatments are performed digitally by using the appropriate software.

2.4.3.4 Cylindrical Mirror Analyzer (CMA) versus the Concentric Hemispherical Analyzer (CHA)

From a historical point of view in AES spectroscopy, the cylindrical mirror analyzer (CMA) (Fig. 8) has nearly always been used as the energy dispersive element whereas for XPS experiments a concentric hemispherical analyzer (CHA) (Fig. 9) has been the choice. A reason for this different approach is to be found in the fact that AES was always considered as an imaging or microscopic technique in which the sample area under investigation was excited by a well focused primary beam scanned over the sample surface. It is for this reason that this approach is called

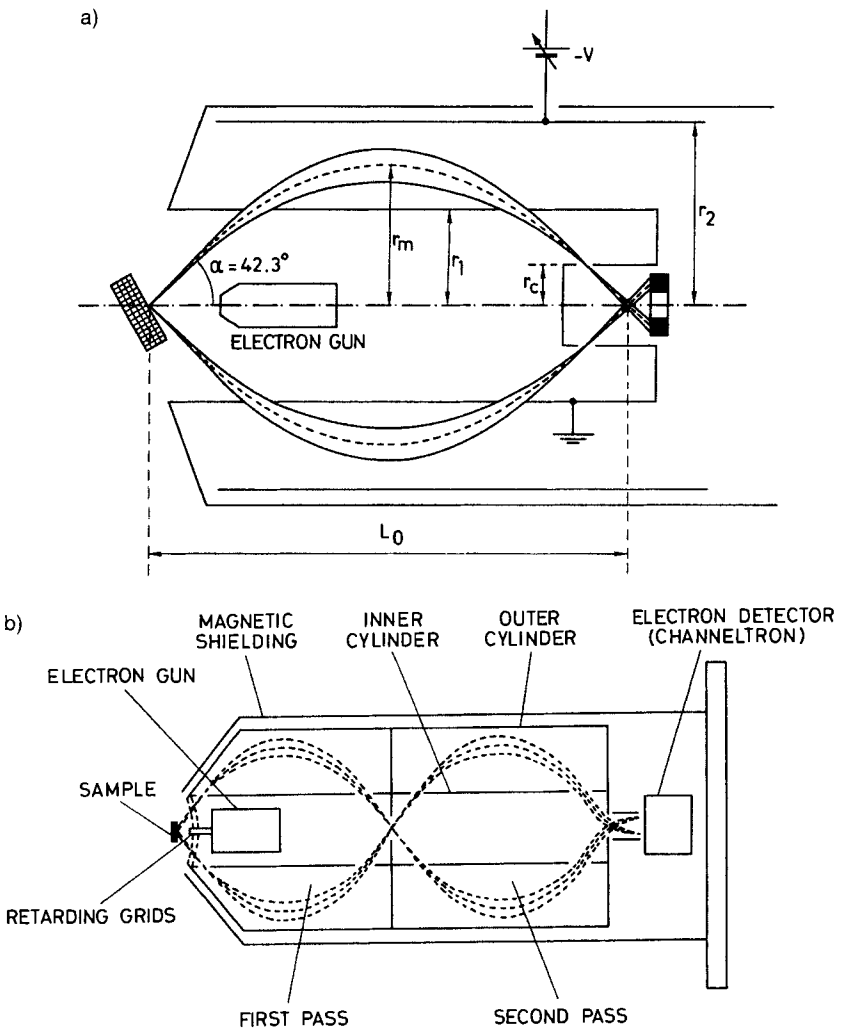


Figure 8. (a) Single-pass CMA; (b) double-pass CMA.

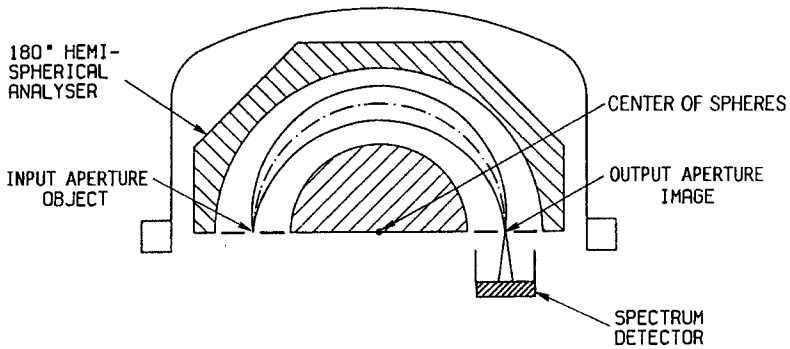


Figure 9. Concentric hemispherical analyzer. Object, image and center of the concentric spheres are on a straight line. This is a general property and not only holds for the 180° device. Adapted from E. Adem, R. Champaneria and P. Coxon, VG Scientific Limited, *Vacuum* **1990,** *41*, 1695 (Fig. 1).

scanning Auger microscopy (SAM). The smaller the excited area, the better the lateral resolution [33–36]. To collect as many emitted Auger electrons as possible from this small excited area, the important parameter is the solid angle Ω under which the electrons are emitted. Under these conditions and for comparable physical dimensions, the CMA was the better choice compared with CHA. However, in XPS experiments in which the primary excitation is due to X-rays, which are much more difficult to focus, the excited sample area is much larger as compared to AES. The electrons to be energy analyzed are no longer collected from a point source, but from a much extended source. This means that no longer is Ω the important parameter but instead the étendue $\lambda = \Omega S$, that is, the product of solid angle of acceptance and area seen by the energy dispersive system. Under these operating conditions a CHA has better performances than the CMA.

The Focusing Properties of the CMA and CHA

The CMA, as shown in Fig. 8(a) and consisting of two concentric cylinders, has become very popular since Zashkvara and coworkers [37] showed that it has unexpected second order focusing properties for $\alpha = 42.3°$ and that refocusing of an object situated at A occurs at B for

$$L_0 = 6.1\, r_1 \tag{25}$$

if

$$\frac{E_0}{qV} = \frac{1.3}{\ln(r_2/r_1)} \tag{26}$$

where E_0 is the energy of the transmitted electrons, and V the potential difference between the two concentric cylinders with radii r_1 and r_2.

Second order focusing means that in the Taylor expansion of the function

$$\Delta L = f(\Delta E, \Delta\alpha) \tag{27}$$

with $\Delta L = L - L_0$, $\Delta E = E - E_0$, $\Delta\alpha = \alpha - \alpha_0$, neither first nor second order terms in $\Delta\alpha$ appear. For a CMA with $\alpha_0 = 42.3°$

$$\frac{\Delta L}{r_1} = 5.6\left(\frac{\Delta E}{E_0}\right) - 15.4(\Delta\alpha)^3 + \ldots \tag{28}$$

which is a figure of merit of the focusing properties of the analyzer as a function of energy resolution $\Delta E/E_0$ and half opening angle $\Delta\alpha$. The influence of $\Delta\alpha$ on the

focusing properties is also called spherical aberration.

In their comparison of the CHA and CMA devices, Hafner et al. [38] showed that a similar Taylor expansion for the 180° CHA has a second order term in $(\Delta\alpha)^2$ but no term in $(\Delta\alpha)^3$. Despite the absence of $(\Delta\alpha)^3$ CHA only has first order focusing. Due to the different order of focusing, a comparison of the performance between the CMA and CHA as a function of $\Delta\alpha$ using the 'reduced dispersion', which is the ratio between the numerical values of $\Delta E/E_0$ to the first significant aberration term in $\Delta\alpha$, is impossible. For this reason, Hafner used the 'reduced aberration', being the ratio between the trace width (a measure of the spatial extension of the beam) and the relative resolution or dispersion, as a figure of merit.

It turns out that despite the second order focusing (not to be confused with double focusing which is used for instance in mass spectrometers of the Mattauch–Herzog type where ions are focused to a single point independently of the angle under which they are injected (geometrical focusing) and irrespective of their energy (energy focusing)) the CMA and CHA have comparable performances. However, as the CMA has a third order aberration term it is known from optics [39] that there is a region different from the image plane where the trace width is at a minimum. For a CMA, this region lies ahead of the image, which is found on the symmetry axis of the instrument, and takes for reasons of symmetry the form of a cylinder, with radius r_c (Fig. 8a), concentric with the deflecting cylinders. If the output aperture coincides with this fictitious cylinder, clearly the energy resolution is greatly enhanced as a smaller aperture can be used, while retaining the same transmission. Under these operating conditions a CMA performs clearly better than a CHA when using the reduced aberration as a criterion. Because of its excellent performance, the CMA has been very popular as an energy dispersive device for both AES and XPS. Today, it is very successful in AES, wherein a very compact instrument can be built if the electron gun is fitted into the inner cylinder of the CMA.

The CMA as used in XPS

For XPS purposes, where traditionally the absolute energy spread ΔE is of prime importance, retardation of incoming electrons to a fixed low pass energy E_0 of the analyzer has become standard practice. The retardation is mostly obtained between hemispherical grids centered around the source area. However, grids suffer from several drawbacks such as scattering of the electrons, generation of secondary electrons strange to the experiment and responsible for extra background noise, and degradation of the angular properties of the beam due to lens action of each mesh of the grid. As an alternative to using grids, other solutions have been proposed such as special input lenses in front of the CMA [40]. However, because the accepted beam in a CMA is confined between two conical surfaces, the design of a retarding lens is not straightforward and, to our knowledge, this idea is not used in commercial equipment.

A major drawback of the CMA is its sensitivity to sample position, both in the axial direction and in the direction perpendicular to the axis. A misalignment of the sample would give a reduction in

transmission, a change in pass energy and a worsening of the resolution. These effects can be greatly reduced by using either a double pass CMA (Fig. 8b) or a single CMA combined downstream with a small 90° hemispherical analyzer. The output aperture of the first CMA acts as the input aperture of the second energy dispersive device and the pass energy of this energy selector no longer depends on the sample position. The double pass [41] scheme also enables the signal-to-noise ratio to be enhanced at the detector while retaining maximum luminosity L (luminosity is the product of the étendue λ with the transmission T, the latter being the ratio of emergent particle flux to incoming particle flux). This approach also facilitates the detection of the energy-filtered electrons because a simple electron multiplier or channeltron can be used instead of the more complex annular detector often recommended in a single pass CMA.

Parallel Detection

When using energy dispersive devices of the focusing type, such as the CMA or CHA, a real image is formed of the input aperture at the exit of the analyzer, which is spatially shifted in the energy dispersive direction depending on the energy under consideration. An energy analysis of the incoming particles can be obtained in the following two ways. Either one uses an exit aperture, which defines the resolution of the analyzer, and onto which the entrance aperture is imaged for one particular energy. By changing the field between the deflecting plates, the pass energy or the energy of the particles passing through the exit aperture, and which hit a detector, can be varied. By sweeping this pass energy, the energy distribution of the incoming particle flux is probed sequentially. Or, one can eliminate the exit aperture and replace it by several detectors positioned in the dispersive direction of the analyzer. In the same manner as an exit slit, the physical dimension of each detector defines an energy spread ΔE and, by using several detectors simultaneously, a region of the energy spectrum can be measured in parallel. Of course, the total energy region probed simultaneously by the different detectors is rather restricted and depends essentially on the physical dimensions of the analyzer and the mean energy transmitted (mean energy has the same meaning as pass energy in an analyzer with an exit aperture). Nevertheless, it is clear that under critical circumstances (i.e., bad signal-to-noise ratios) the principle of parallel detection alleviates the noise problem and reduces the measuring time inversely according to the number of detectors.

In modern equipment for AES and XPS the concept of parallel detection is fully implemented for many applications. As in a CHA, the energy dispersive direction coincides with the produced direction of the input aperture and the center of the deflecting spheres. There is also good accessibility for mounting several detectors. In a CMA, however, where the energy dispersive direction coincides with the symmetry axis of the device, parallel detection is not as easily implemented as in a CHA.

Combination CHA-Input or -Transfer Lens

As already outlined, CMA based systems have traditionally been used for small spot work such as with AES, whereas CHA

PAGE MISSING OR DUPLLICATE

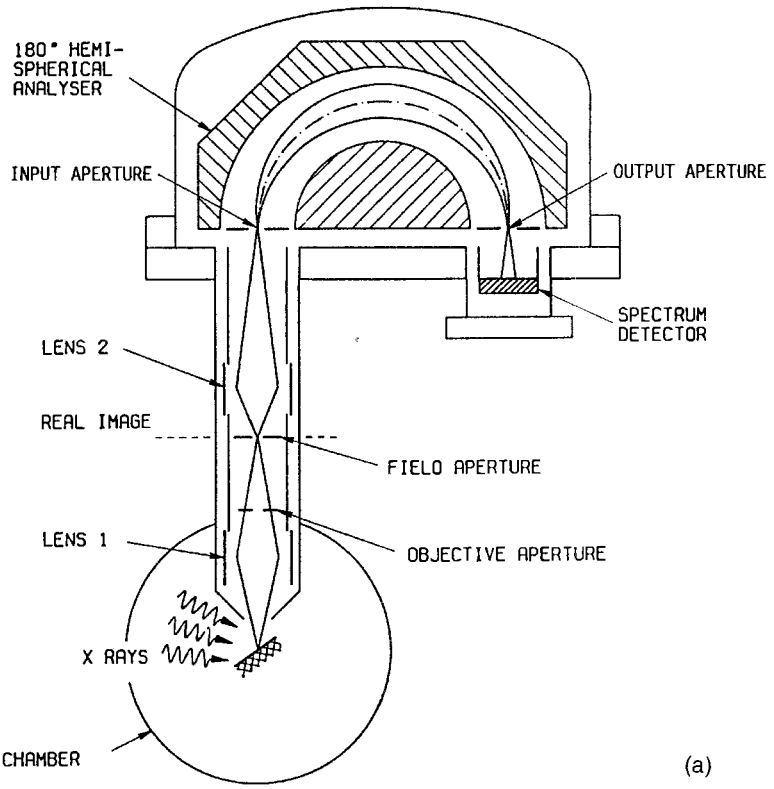


Figure 10. (a) Combination of transfer lens and CHA defining the field of view and the acceptance angle.

kind of magnetic lens close to the sample with the ***B*** vector pointing towards the transfer lens.

The situation becomes even more complex during XPS experiments when flooding electrons are required. Simple straight-on flood guns cannot be used as they illuminate the sample under a direction different from the axis of the transfer lens or ***B*** vector and special layouts, coaxial with the transfer lens or ***B*** vector, are needed.

Generally, the purpose of the electrostatic transfer lens is threefold [45]. Namely, (i) it defines the field of view, (ii) it defines the acceptance angle, and (iii) it (possibly) retards the incoming electrons in the CHA. Figure 10a shows an input or transfer lens in front of the CHA which, as can be seen, comprises two separate lenses. The sample surface is situated in the object plane of lens 1 and is imaged on the field aperture situated in the image plane of the same lens: that is, the sample and field aperture are conjugate. The area seen on the sample or the field of view is defined by the aperture area of the field aperture and by the magnification of the lens, depending on the potentials applied to the lens elements. The field of view area is simply the ratio of the area of the field aperture to the lens magnification.

The objective aperture, preceding of the field aperture, defines the acceptance angle of the incoming beam. The real image sharpness (or spatial resolution) at

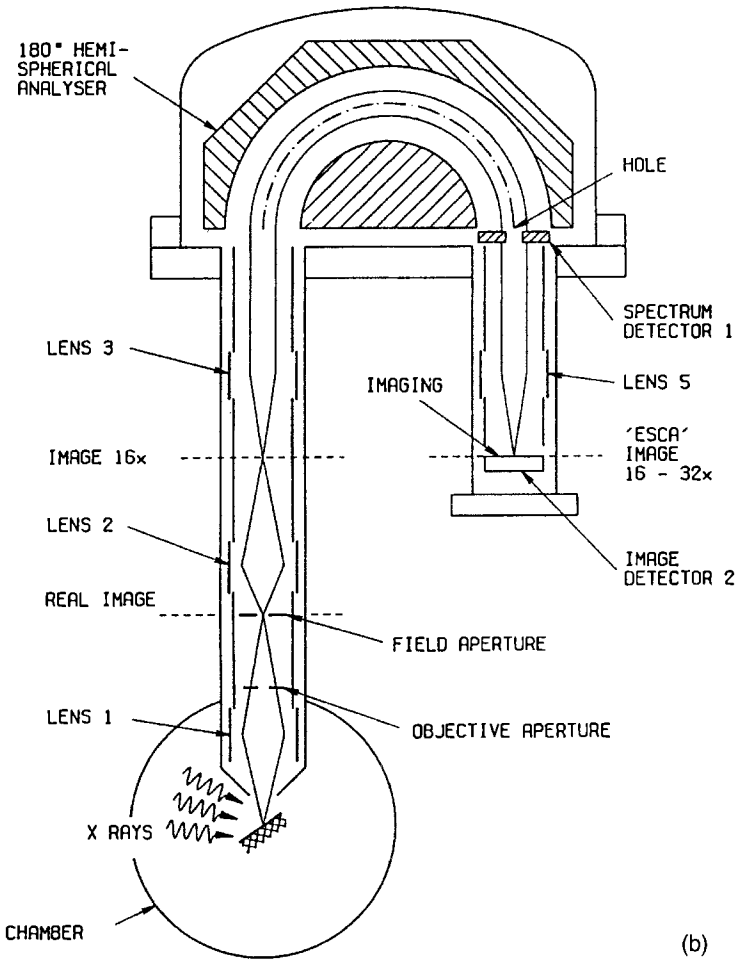


Figure 10 (b) Similar setup as in (a), but with additional Fourier lenses 3 and 5. Adapted from E. Adem, R. Champaneria and P. Coxon, VG Scientific Limited, *Vacuum* **1990,** *41*, 1695. Reproduced by permission of P. Coxon et al. and Pergamon Press.

the field aperture, which depends on lens aberrations, is enhanced by reducing the area of the objective aperture, just as in classical optics. However, the better the sharpness, the smaller the intensity or the number of electrons emitted from the field of view passing through the lens.

The second lens forms an image of the field aperture on the input aperture of the CHA. This lens provides extra magnification and also performs the decelerating function (without the use of grids) if the CHA is used in constant pass energy mode. The degree of lens filling (i.e., the radial trajectory distribution) depends on the amount of deceleration which means that the spherical aberration of this lens also depends on the degree of retardation. The complete behavior of such a lens is best understood by trajectory simulations [46]. To reiterate: the deceleration just before entering the CHA vastly enhances the performance of the CHA combined with a transfer lens and thus makes this combination particularly attractive for XPS.

2.4.3.5 Imaging Techniques

As already discussed, there are two distinct methods used in AES/XPS imaging: the source defined (SD) approach and the detector defined (DD) approach. In the SD mode, lateral resolution is obtained by good focusing of the primary excitation source. In AES, focusing presents no problem since it originates from the field of scanning electron microscopy.

Imaging AES or Scanning Auger Microscopy

A scanning Auger microscope (SAM) is, in principle, a combination of a scanning electron microscope (SEM) and an Auger electron energy analyzer (CMA or CHA) [47]. Indeed, in a SEM a focused beam of relatively high energy (25–65 keV) electrons not only produces the backscattered electrons (BSE) and the true secondary electrons (SE) with which SEM images are formed, but also Auger electrons at the point of impact. Taking into account the limited escape depth of Auger electrons, only the uppermost layers contribute to the signal, although the penetration depth of the primary beam may be of the order of 1 μm.

When an energy analyzer is introduced in such an instrument and calibrated on a fixed energy corresponding to a particular Auger transition, a composition image of the surface layer can be formed with a lateral resolution equal to or even better than the resolution of the SEM image. Indeed, the resolution of the SEM image is somewhat lower than the beam diameter as a result of backscattering electrons emerging from the specimen relatively far from the point of impact. It has been shown, however, that Auger electrons emerge practically exclusively from the point of impact; therefore, the resolution is in principle given by the beam diameter [48]. In the commercially available dedicated SAM instruments the primary energy of the electron beam is limited to 10 keV to increase Auger electron production. Indeed, since the ionization cross section for Auger electron emission decreases when using higher primary energies, the lateral resolution is lowered by a considerable amount, in most instruments by around 100 nm.

The deflection plates of a conventional SAM are added to the electron gun optics to be able to raster the surface of the sample. In general, lateral resolutions sufficient to study problems in metallurgy or microelectronics are readily obtained and spatial distributions of elements in a surface can be studied. In principle, it is also possible to map a component of a composite Auger line when the element under examination is present in different chemical environments, as discussed above. The SAM has found widespread use in solving industrial problems on samples which are microscopically nonhomogeneous and sometimes exhibiting rather rough surfaces. Quantification of the surface concentrations on such samples has to be undertaken with great care, as was described by Prutton et al. from the York group [49]. We will return to these problems in more detail in what follows.

In most commercial SAM instruments a secondary electron detector (SED) is included to produce SEM images of the surface under study. Furthermore, absorption current images can also be obtained. Both are very useful in selecting the surface area which has to be studied with the SAM.

The higher Auger production efficiency, when working at lower primary energies, allows a reduction of the primary beam current (~10 nA) and consequently limits the beam effects. Additional to two-dimensional imaging of the surface chemical composition, called element mapping, line scans can also be used in combination with SEM images, producing SEM images combined with elemental composition lateral profiles.

Mapping in scanning Auger microscopy is a time-consuming procedure. A reduction in measuring time can be obtained by using electron emitting sources with a brightness as high as possible. In this respect the introduction of LaB_6 emitters has proven to be a leap forward in brightness as compared with the classic thermionic emitters. However, LaB_6 sources suffer from two drawbacks. First, their brightness, in other words the amount of current extracted from the filament tip per unit area and solid angle, decreases drastically when, for instance, hydrocarbons which can react with the LaB_6 filament are in the vicinity. In this respect, several oxygen treatments have been proposed which enhance the performance of the filament although there remains a degradation in brightness and in emitting stability. Second, the emitting tip of the LaB_6 filament is rather sensitive to ion backsputtering. As soon as the pressure in the neighborhood of the filament rises above 10^{-6} mbar there is sufficient restgas ionization to produce positive ions which are accelerated towards the filament and sputter the filament. This sputtering degrades the geometry of the electron emitting tip, decreasing the geometrical quality of the electron source.

The stability problems of LaB_6 emitters are largely shelved by using a new generation of electron emitters, namely, Schottky field emitters. These consist of a single crystal tungsten tip coated with a zirconia layer to lower the work function of the tip. Typical dimensions of the electron emitting surface are of the order of 250 nm. Compared with LaB_6 emitters, Schottky field emitters have higher brightness, lower energy spreads and longer lifetimes. However, due to the needle shape of the emitter, backsputtering must again be eliminated completely requiring a dedicated pumping system in the vicinity of the Schottky emitter.

Figure 11 shows an example of an SAM image recorded using a Schottky electron emitter. The image was recorded using the Fisons Microlab 310-F field emission scanning Auger microprobe. It shows a 15 nm wide SiO_2 gate oxide layer, 200 nm below

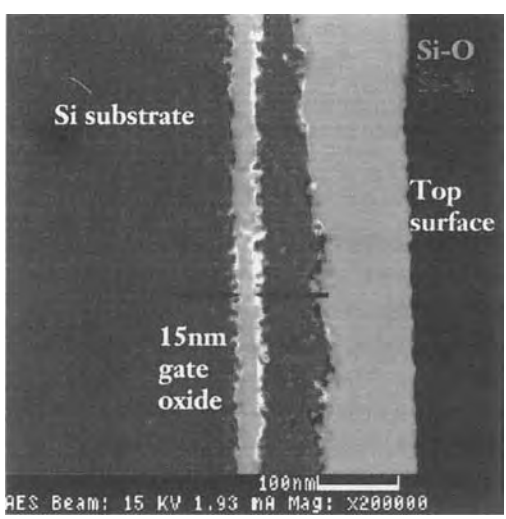


Figure 11. A 15-nm wide SiO_2 gate oxide layer, 200 nm below the top surface of an Si wafer, as revealed using the Fisons Microlab SAM on a fractured edge of the wafer. Reproduced by permission of Fisons Instruments.

the top surface of an Si wafer. The image was obtained by recording the Si–Auger signal characteristic, respectively, for the bulk material and for the oxide.

Imaging XPS

In scanning Auger microscopy, at electron energies of ~2 keV, resolutions of the order of 50 nm are attainable with beam currents of the order of 10^{-9} A. This corresponds to current densities of 40 A cm^{-2} and pinpoints one of the weaknesses of AES or SAM, namely sample damage. In XPS, the damage due to X-rays is much lower; however, it is much more difficult to focus X-rays.

Generation of X-rays: X-rays are classically generated by energetic electrons impinging on an anode or anticathode. The type of anticathode material defines the X-ray spectrum. Besides the characteristic lines, background bremsstrahlung is also generated although this can be reduced considerably by using appropriate filtering window materials. Nevertheless, bremsstrahlung levels of as high as 20% of the principal characteristic emission lines remain.

With the simple cathode–anticathode arrangement the emerging X-ray beam is usually very broad and so spatial resolution can only be obtained by collimating the beam. This is done at the expense of intensity. Besides intensity, there remain several other drawbacks and limitations for XPS work using a simple X-ray source. For high energy resolution XPS, there is often a lack of X-ray line sharpness due to satellite and ghost lines [50] which can complicate the measured XPS spectrum. Also, there is the remaining continuum radiation which, on the one hand, can damage sensitive samples while, on the other hand, generating extra noise in the XPS detection equipment. A complete survey of these effects is given by Chaney [51].

Focusing and monochromatizing the X-rays: The benefits of using a crystal monochromator to eliminate the aforementioned problems has been discussed by Siegbahn [52] and coworkers. If a crystal lattice is bent as shown in Fig. 12a, with a radius equal to the diameter of the Rowland circle, X-rays originating from a point source on the Rowland circle can be diffracted and monochromatized from an extended area of the bent crystal, thus collecting X-rays in an opening angle proportional to the diffracting area. The Bragg relation,

$$n\lambda = 2d \sin\theta \tag{29}$$

with usual notation, is satisfied along the whole diffracting area. As the X-rays are also reflected under specular conditions, refocusing of the monochromatized X-rays also occurs at a point on the Rowland circle. Rowland first recognized the imaging properties of this setup for a spherical grating and it was Johann who first implemented these ideas in a practical arrangement for X-rays by bending a crystal lattice. As is clear from Fig. 12a, some focusing errors remain. These were corrected by Johansson in the layout shown in Fig. 12b. Johansson not only bent the crystal, but ground it to the correct shape. The practical implementation of this approach is named after Johansson or Guinier. Focusing X-ray monochromators such as these became a standard tool in XPS and are used in UHV

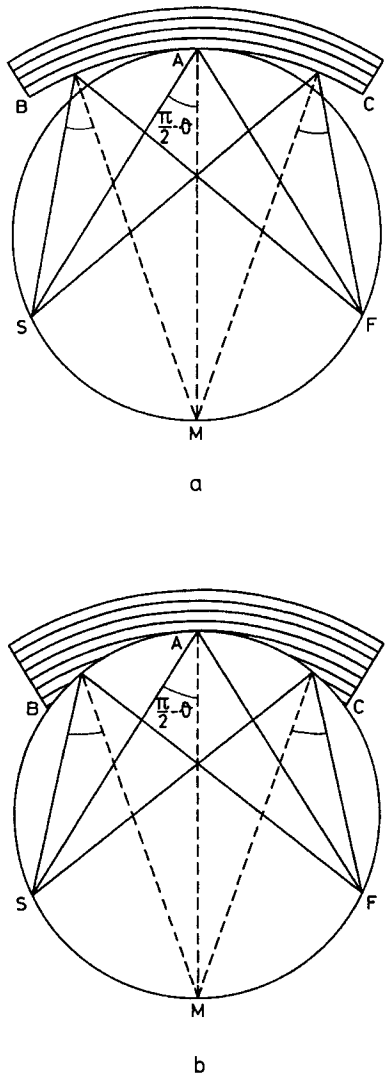


Figure 12. (a) The Johann solution: the bent crystal has a radius equal to the diameter of the Rowland circle; (b) the Johansson or Guinier solution.

apparatus as shown in Fig. 13. In practice, several bent quartz crystals are used to monochromatize and refocus as much X-ray flux as possible. To reduce astigmatism, the crystals are bent toroidally, not spherically. With this equipment, it is possible to obtain illuminated spot sizes of the order of 50–100 μm.

'Source defined' imaging in XPS: With these dimensions as lateral resolution, some kind of source defined (SD) imaging work already becomes possible and the trivial answer to providing a sequential pixeling is performed by moving the sample stage with a precision X–Y manipulator, storing the XPS energy information pixel by pixel. This crude approach is, of course, extremely time consuming and certainly not practical for extensive use.

A promising approach to SD XPS imaging was given early on by Hovland [53]. This author coated the back of his thin samples with an aluminum layer which fulfilled the function of anticathode. A scanning electron beam impinging on the back of his sample, on the aluminum layer, generated a scanning X-ray beam, exciting the sample and producing the photoelectrons. However, due to the need for elaborate sample preparation and, in any case, unsatisfactory results, this idea was abandoned.

A breakthrough in the pixeling process has been made by Physical Electronics Inc. using elliptically bent crystals. The principle underlying this approach is shown in Fig. 14. At the foci of the ellipse, which defines the bent crystals lattice surface, are the anticathode and sample. By scanning the X-ray generating electron beam over the anticathode, this anticathode is sequentially pixeled. Each pixel emits X-rays which are monochromatized and reflected on the sample surface by the elliptically bent lattice. In other words, each electron beam defined pixel at the anticathode is X-ray imaged at the sample and by scanning the electron beam over the anticathode, a monochromatized and focused X-ray beam scanning the sample is obtained. This provides an elegant

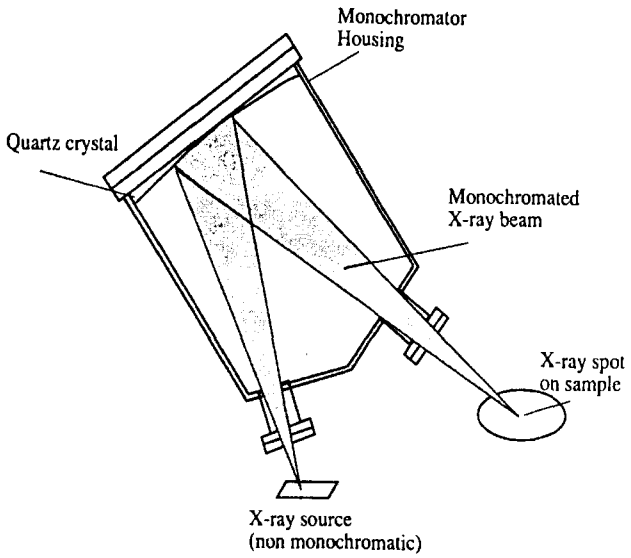


Figure 13. Practical focusing X-ray monochromator of the Johansson or Guinier configuration, ultrahigh vacuum adapted for XPS work. Adapted from Kratos Analytical, Axis product brief. Reproduced by permission of Kratos Analytical Ltd.

solution to SD XPS imaging. In principle the lateral dimensions of the X-ray spot are of the same order of magnitude as the dimensions of the electron beam on the anticathode. For XPS imaging, this leads to lateral resolutions of the order of 10 μm. Besides imaging, this system enables point mode measurements to be made (i.e., enables more standard XPS measurements to be made by keeping the small spot X-ray beam fixed on a particular position on the sample).

'Detector defined' imaging in XPS: This aforementioned and elegant solution to source defined XPS imaging strongly contrasts with the early attempts of detector defined (DD) XPS imaging in which an extended area of the sample is illuminated with X-rays and, by preference, also monochromatized. In the first attempts, Keast and Dowing [54] simply inserted a collimator between the sample and their energy dispersive system to achieve some degree of lateral resolution and again, by

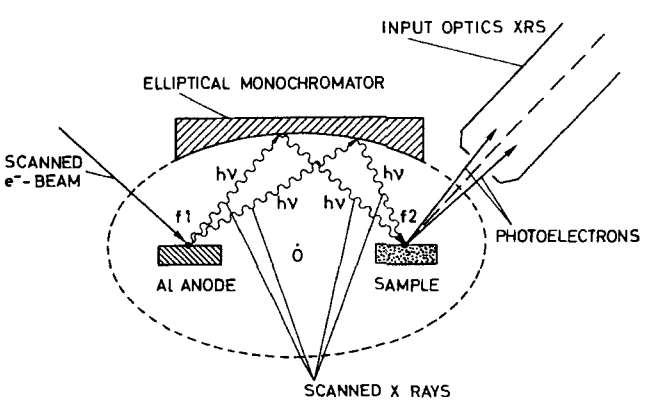


Figure 14. A scanning electron beam on the Al-anticathode is transformed, with the aid of an elliptically bent crystal, into a monochromatized scanning X-ray beam focused on the sample. Adapted from Physical Electronics Quantum 2000 documentation. Reproduced by permission of Physical Electronics Inc.

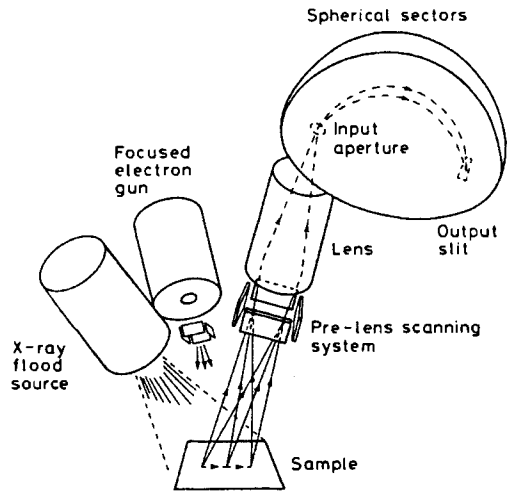


Figure 15. In the detection defined XPS imaging, pixeling can be obtained by the use of a pre-lens scanning system. Adapted from M. P. Seah and G. C. Smith, *Surf. Interface Anal.* **1988,** *11,* 69 (Fig. 3). Reproduced by permission of M. P. Seah and John Wiley & Sons Ltd.

moving the sample stage, a time-consuming and trivial solution to detection defined XPS imaging can be attained.

A second logical step in DD XPS imaging has already been mentioned. Namely, this approach entails pixeling by using scanning deflection plates (Fig. 15) in front of the transfer lens of the CHA [55]. The transfer lens defines the field of view or the lateral resolution, whereas the scan plates provide for the pixeling by sequentially collecting the emitted electrons from the different virtual pixels. In this approach, only those electrons emitted from the pixel addressed by the scan plates are detected, while electrons emitted from other areas of the X-ray illuminated sample are not accepted by the combination of transfer lens and CHA. This is clearly a waste of information and translates itself into longer acquisition times for building up an image of sufficient quality.

To improve the detection efficiency in the DD imaging mode, Gurker et al. [56, 57] exploited the imaging properties of the CHA in combination with a two-dimensional position sensitive detector. As previously discussed, a CHA is a first order imaging device in two orthogonal directions. As a consequence, a point source at the entrance plane is imaged in the output plane as a spot with the same dimensions as the source. The object plane (input) and image plane (output) are conjugate. This is illustrated in Fig. 16 in which electrons are emitted from the same spot on the sample under three different energy conditions. When all electrons have the same energy, and if the pass energy of

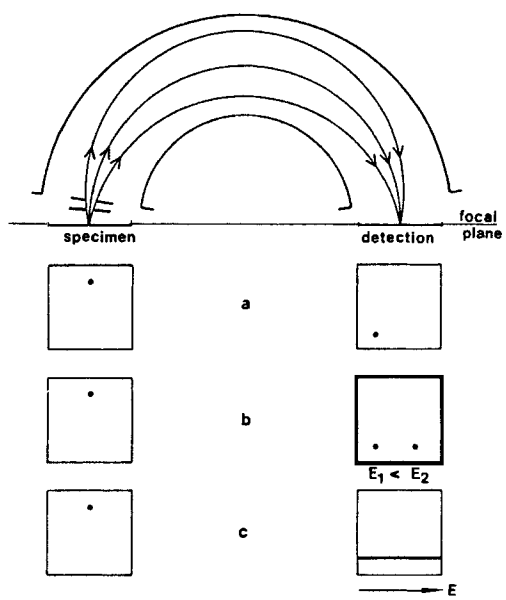


Figure 16. Imaging properties of the CHA as a function of energy. (a) All electrons have the same energy; (b) electrons with two different energies E_1 and E_2 are emitted from the same spot; (c) electrons with a continuous energy distribution are emitted from the same source spot. Adapted from N. Gurker et al., *Surf. Interface Anal.* **1987,** *10,* 243 (Fig. 1). Reproduced by permission of N. Gurker and John Wiley & Sons Ltd.

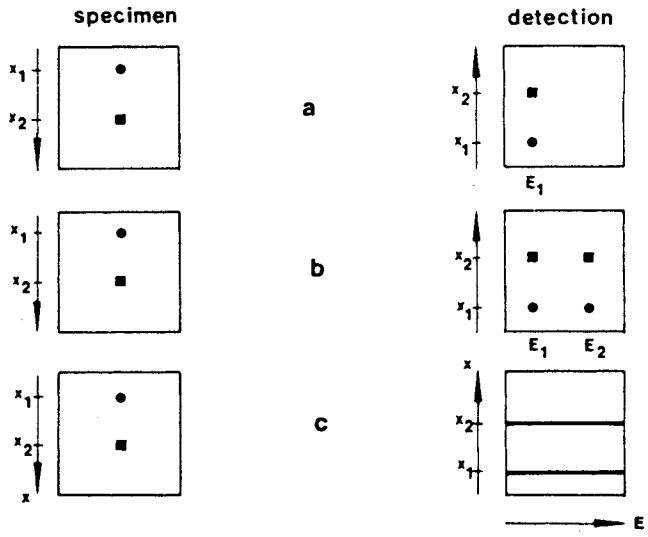


Figure 17. Source displacement in the x direction has no influence on the image displacement, due to energy dispersion which is perpendicular to the x direction. Adapted from N. Gurker et al., *Surf. Interface Anal.* **1987**, *10*, 243 (Fig. 2). Reproduced by permission of N. Gurker and John Wiley & Sons Ltd.

the CHA is tuned close enough to that energy, these electrons are imaged as a single spot in the detection plane as is shown in situation (a). If, however, electrons with two different energies, situation (b), are emitted from the same spot, they will refocus in the detection plane at two different spot positions, depending on their respective energies. More generally, if electrons with a continuous energy distribution are emitted from the sample spot, they will refocus following a straight line in the detection plane. As shown in situation (c), this line is parallel with the energy dispersive direction of the CHA. Figure 17 shows the behavior of the system if electrons are emitted from two different spot positions. The spot positions on the sample are chosen such that they follow the x direction, that is, perpendicular to the energy dispersive directions of the CHA. In this way, the imaging function of the CHA is not convoluted with the energy dispersing function of the CHA and both kinds of information remain simultaneously accessible on a two-dimensional position sensitive detector. In other words, by using a slit as aperture at the entrance, perpendicular to the energy dispersive direction of the CHA, a strip-like 'field of view' at the sample is projected on the detector again as a strip-like image. The one-dimensional spatial information is displayed following a direction perpendicular to the energy dispersive direction while the energy information is accessible in the parallel direction. Without yielding true two-dimensional imaging, this setup fully exploits the possibilities of the CHA combined with a two-dimensional position sensitive detector, because of the parallel detection scheme. Complete two-dimensional imaging can be achieved by moving the sample stage in the energy dispersive direction and taking successive linescans to produce a true XPS image at a fixed energy. A similar approach has been adopted by Allison [58] and Scienta Instrument A.B.

Recently, VG Scientific Ltd have pioneered a breakthrough on the side of DD

XPS imaging by fully exploiting the possibilities and potentialities of their transfer lens system combined with an extra lens at the exit of the CHA [59, 60]. Figure 10b gives a schematic picture of this lens combination. Compared with Fig. 10a, the only difference with the previously described transfer lens–CHA system is to be found in the extra 'Fourier' lenses 3 and 5. The Fourier lens gets its name from the fact that the real image of the sample produced by lenses 1 and 2 is situated in the focal plane of the Fourier lens 3. Similarly, the image detector, behind of the complete electron optical system, is in the focal plane of lens 5. As is well known from optics, a point source situated in the focal plane of a lens produces a parallel beam of light downstream of the lens and vice versa; exactly the same happens with the Fourier lenses of Fig. 10b. Electrons emitted from a single point in the image plane of lens 2, which coincides with the focal plane of Fourier lens 3, give rise to a parallel beam of electrons through the CHA. Similarly, a parallel beam of electrons entering Fourier lens 5 are refocused to a single point at the focal plane of lens 5 on to the image detector.

To understand how imaging is combined with energy selection we refer to Fig. 18a. For simplicity, the transfer lens consists of two idealized thin lenses: that is, a first lens combining the function of lens 1 and lens 2 in Fig. 10b, magnifying, defining the field of view, and the degree of lens filling or acceptance; and, a second lens acting as a Fourier lens. In other words, the field of view aperture of the first lens is the real object of Fourier lens 2 and is situated at the focal plane of this Fourier lens. As previously described, the field of view aperture in the image plane of lens 1 defines the field of view on the sample. Narrowing the field of view aperture reduces the field of view on the sample. From each edge of this field of view, all electrons are of course refocused at the edge of the field of view aperture but only those electrons within the hatched beam of Fig. 18a can be transmitted through aperture 3 of the CHA as a parallel beam of electrons. Narrowing this input aperture 3 of the analyzer does not change the field of view but only narrows the opening angle of the hatched beam at the sample, again at the expense of intensity. The same reasoning holds for each point of the sample within the field of view: in other words, each point emits electrons in an opening angle defined by aperture 3 and produces a parallel beam of electrons into the CHA under a well-defined angle between the limiting angles of Fig. 18a corresponding with the edges of the field of view. Two important things are to be remembered: first, a specific point within the field of view is translated into a specific beam direction into the CHA; second, by reducing the size of aperture 3 no image information is lost, only intensity!

Figure 18b gives a simple overview of this transfer lens behavior. The double ended arrow on the sample defines the field of view and is transformed or magnified into a real image at the aperture field of view. Each point of this image is transformed into a particular beam direction in the CHA. As a 180° CHA is a 1 : 1 imaging device, all beam directions at the input aperture are transformed into similar beam directions at the output aperture. The input aperture of the CHA is correctly imaged on to the output aperture only for those electron energies at the pass energy

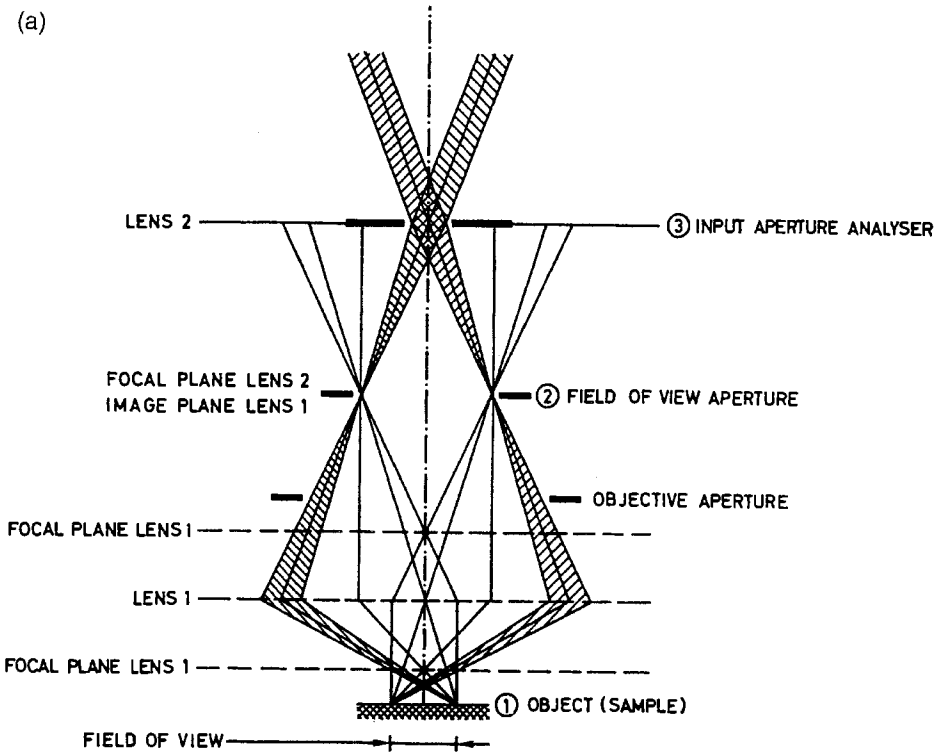


Figure 18. (a) Simplified picture of the functioning of the single input transfer lens 1 combined with the Fourier lens 2

E_0, giving the required energy dispersion. The energy resolution can be changed by changing the size of output and input aperture but, as discussed above, this has no influence on the image information, only on the intensity which translates itself into acquisition time of the image.

Finally, the Fourier lens at the output of the CHA transforms all directions back to a real image on the imaging detector, but at a well-defined energy.

Figure 19 shows an example of an XPS image, recorded using the VG Scientific Ltd ESCALAB 220i fitted with the XL lens. It shows the $W4f_{7/2}$ image of 5 μm W bars with 5 μm spacing, on an Si substrate. The resolution is better than 2 μm.

2.4.3.6 Magnetic Fields in Imaging XPS

The idea of using magnetic fields for XPS imaging is not new. Already in the early 1980s, Beamson and coworkers [61, 62] proposed a true XPS microscope in which a superconducting seven tesla magnet was used. Similar equipment, not typically for XPS purposes, has been developed with electromagnetics [63] and even with permanent magnets [64]. In these special arrangements nonhomogeneous magnetic fields [65] are essential. In the high field region each emitted electron spirals around a magnetic field line under

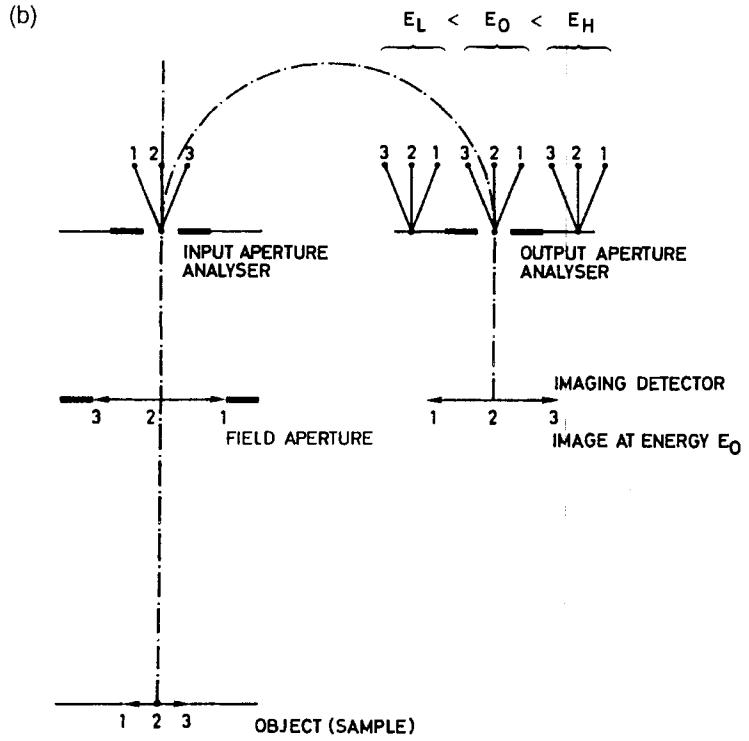


Figure 18 (b) simplified overall picture of the imaging process as obtained with the aid of Fourier lenses at the input and output of the CHA.

the influence of a Lorentz force. If the change in magnetic field strength during one such spiraling motion is compared with the average magnetic field, the total energy and angular momentum are conserved quantities (adiabatic motion) [63]. The total conserved energy of an electron in such a magnetic field is the sum of an axial component, along the field lines, and a cyclotron component. Due to conservation of angular momentum, this cyclotron component of the electron energy decreases as the field strength decreases in the nonhomogeneous diverging magnetic field. The net effect is an increase in the axial component at the expense of the cyclotron energy component and results, as is shown in Fig. 20, in unspiraling and parallelization of the electron trajectories initially collected in a 2π steradian solid angle in the high field zone. With this approach, Beamson obtained a true photoelectron microscope with very fast image acquisition and acceptable spatial resolution, but without any

Figure 19. 5 μm wide W bars, with 5 μm spacing, on an Si substrate. Reproduced by permission of VG Scientific Ltd.

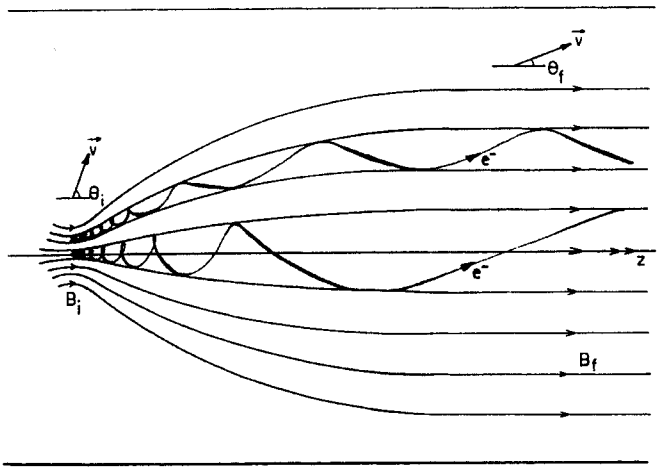


Figure 20. Electrons emitted in the high field region of an inhomogeneous divergent magnetic field are parallelized and unspiraled in the low field region. Adapted from P. Kruit and F. H. Read, *J. Phys. E* **1983,** *16*, 313 (Fig. 1). Reproduced by permission of N. Kruit and The Institute of Physics (Great Britain).

inherent energy dispersive capability. By inserting a retarding field, high pass energy images could be obtained, but no true energy dispersive imaging has been reported. Although very promising, no widespread use has been made of such inhomogeneous magnetic fields for XPS imaging. This is probably due to several reasons. As already mentioned, with magnetic fields close to the sample, it is very difficult to provide for instance flooding electrons used to avoid charging of the sample. The intense magnetic fields as used in the Beamson approach will also compromise the correct functioning of any type of electrostatic energy analyzer used to obtain energy-selected images.

A much simpler, yet efficient, approach in XPS imaging relies on the use of a magnetic objective lens, situated as close to the sample as possible (Fig. 21). These lenses do not provide an acceptance angle as large as with the previously discussed nonhomogeneous magnetic fields. However, compared with electrostatic objective lenses, they constitute a much improved acceptance angle. Due to the low aberrations of magnetic lenses, this increase in acceptance is obtained without loss of spatial or energy resolution. Of course this improvement in an opening angle is translated into a net reduction in acquisition time, that is, the time required to build up an energy selected image of good enough quality. The combination of such a magnetic objective lens with a Fourier imaging system and a CHA yields spatial resolutions [66] of the order of 1 μm.

2.4.4 Characteristics of Scanning Auger Microscopy Images

2.4.4.1 General Aspects

The problems encountered in correctly interpreting SAM images have been discussed in considerable detail by Prutton et al. [49]. This discussion summarizes the principal conclusions obtained by these authors. These authors recognize five basic complicating factors which might give rise to image artefacts:

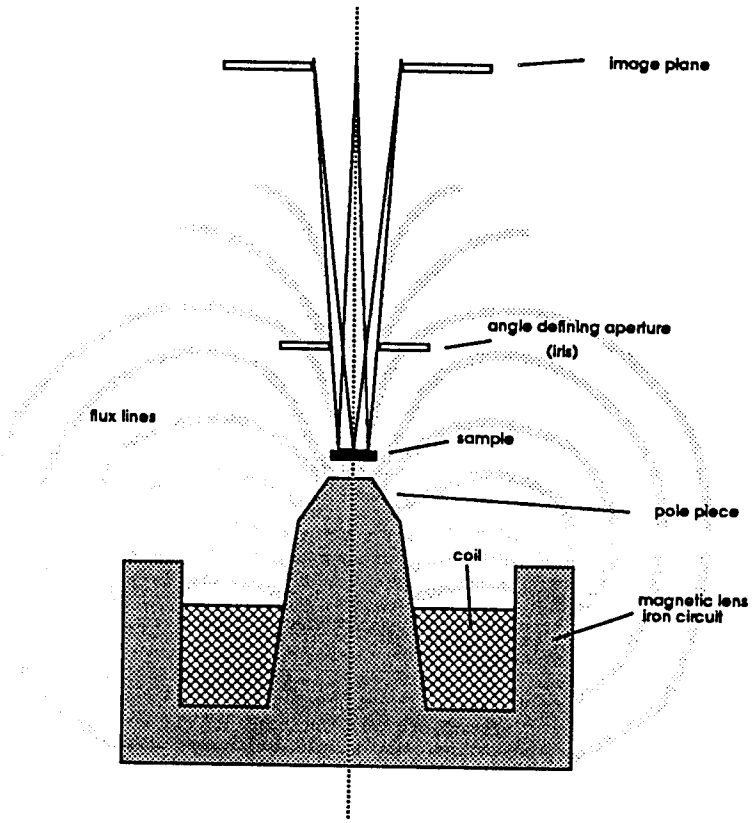


Figure 21. Magnetic objective lens as used in XPS. Adapted from Kratos Axis 165 System Overview. Reproduced by permission of Kratos Analytical Ltd.

(i) Since in SAM imaging the contrast in the Auger image is often obtained from a comparison of peak heights in the Auger peak of an element and in the background, background slope effects could sometimes lead to an apparent Auger signal although the particular element is not even present.

(ii) A uniform layer does not necessarily lead to a uniform Auger image, if the substrate under the layer is nonhomogeneous and the Auger backscattering factor varies from place to place.

(iii) The changes in Auger signal due to surface topography or roughness may be so large that they are able to obscure the variations in chemical composition at the surface.

(iv) In view of the large scan times that are necessary to obtain favorable signal-to-noise and signal-to-beam Auger signal, beam current fluctuations can occur during the measurements, thus spoiling the SAM image contrast.

(v) Extreme surface topographics, such as sharp edges, can lead to erroneous chemical compositions as a result of shadowing and/or enhancement effects on the Auger signal.

Prutton et al. described how these effects, which we will now consider in detail, can be remedied in the MULSAM instrument (multispectral Auger microscope). The results obtained on a sample

consisting of a silicon substrate, a gold overlay, an SiGe alloy on top of part of the gold overlay, and the SiGe substrate are discussed.

This sample exhibits four surface regions:

- SiGe on top of the Si substrate
- SiGe on top of the Au overlay
- the Si substrate
- the Au overlay.

This sample should reveal three different regions when imaging the Si concentration in the surface layer: the brightest corresponding to the Si substrate, followed by the next brightest corresponding to Si, and finally a region without Si when the beam strikes the Au overlay.

2.4.4.2 Background Slope Effects

When the image contrast is formed from the quantity N_1-N_2, where N_1 represents the counts in the peak and N_2 in the background, or if a linearly extrapolated background is used, four regions are revealed instead of three and Si seems to be more concentrated in the Au overlayer, which, of course, is an erroneous result.

This artefact is explained by the background slope in the spectrum when the beam strikes the Au overlayer. Indeed, N_1 and N_2 were measured at 82 and 102 eV, respectively, which is a justifiable choice when measuring the $SiL_{2,3}VV$ Auger peak. The curvature of the background, however, when measuring on Au leads to a larger difference between the signals at 82 and 102 eV than on Si, although no peak is present. From this it is concluded that a more adequate background subtraction procedure is necessary, for instance, by taking more calibration points on the curve. Unfortunately, this increases the measuring time considerably.

2.4.4.3 Substrate Backscattering Effects

A difference in intensity of the Si signal in the region SiGe, respectively, above Si and Au, is observed, although the same Si concentration in the surface layer is present. This is due to a difference in the Auger backscattering factor R for both regions. Prutton et al. [49] described how, using backscattering electron detectors in the MULSAM instrument, allowed an estimation of the Auger backscattering factors in an image obtained simultaneously with the energy-analyzed images. When this correction is carried through a uniform intensity for the SiGe region is obtained and a quantitative analysis of the relative Si concentration gives a result in agreement with the expected value for the alloy SiGe.

2.4.4.4 Topographic Effects

The angle of incidence ϕ, measured relative to the surface normal of the primary electron beam, affects the Auger yield. The number of core ionizations within the Auger escape depth increases as $1/\cos\phi$ on the one hand, and R varies with ϕ since the energy and angular distribution of backscattered electrons is a function of ϕ, on the other hand. Furthermore, the fraction of the total Auger yield collected by the electron spectrometer is

dependent upon the take-off angle θ, also measured relative to the surface normal, because the Auger emission is not isotropic, but exhibits approximately a $\cos\theta$ emission distribution. Therefore, not taking into account obstruction and shadowing or Auger electron diffraction effects, it is expected that the Auger yield will vary as

$$I_A = I_0 \frac{1}{\cos\phi} R(\phi) \cos\theta \qquad (30)$$

Prutton and his colleagues continued by explaining how, in the MULSAM instrument, four BSE detectors placed coaxially around the primary beam, produced difference signals, which enable a topography correction to be made. This was tested on an anisotropically etched (0 0 1) slice and it was shown that when this sample, showing eight (3 3 1) faces, four (1 1 1) faces, and one (1 0 0) face, is covered by a carbon layer of ca 100 nm thickness, almost homogeneous carbon intensity can be obtained. The correction procedure is fully described in their paper [49].

2.4.4.5 Beam Current Fluctuation Effects

Again, Prutton and his colleagues describe how the acquisition of both a secondary electron image and a sample absorption current image is necessary to correct for beam current fluctuations. Both images are anticorrelated, a feature exploited to form a beam current image, which is then divided into any other images obtained at the same time, to correct them to a constant beam current.

2.4.4.6 Edge Effects

Prutton and colleagues also considered the situation of Fig. 22 where the step sidewalls of a metallic overlayer on an Si substrate are shown. On the substrate there will be an obstructed region and a shadowed region (for obvious reasons) and an enhanced region as a result of elastic and inelastic scattering effects. The former can give rise to electrons with a more grazing incidence and the latter to lower primary energy electrons. Both are favorable for increasing the Auger electron yield of the substrate in the region directly adjacent to the metallic overlayer and next to the shadowed region.

These are the well documented edge effects which have to be taken into account when interpreting the Auger images. Prutton et al. [49] explain how this could be achieved and we refer the reader to this paper for more details.

Correction of the different image artefacts, described above, is possible in principle. However, it necessitates a number of supplementary measurements as well as their accompanying software routines. These are not installed on all commercially available instruments and the operator

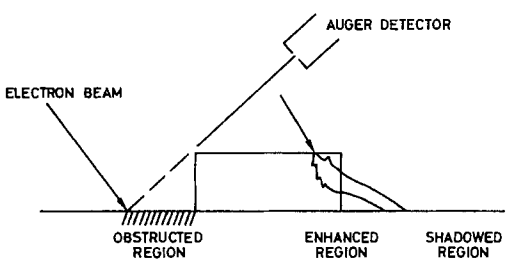


Figure 22. Different regions as discussed in text. Adapted from M. Prutton et al., *Surf. Interface Anal.* **1992,** *18*, 295 (Fig. 3). Reproduced by permission of M. Prutton and John Wiley & Sons Ltd.

should, therefore, be aware of the existence of these artefacts.

2.4.5 Conclusion

Scanning Auger microscopy is already in a well-developed state although imaging XPS is still in its infancy. It is, however, a rapidly evolving field as far as the introduction of new instruments is concerned. Characteristics of XPS images are therefore not considered in detail but case studies are discussed in Volume 2 of this Handbook. In this Chapter only scanning Auger microscopes have been discussed. Combinations of STEM instruments with Auger spectroscopy leads to instruments with impressive lateral resolutions. We refer the reader to the literature for a discussion of these combinations [67].

2.4.6 References

[1] L. Fiermans, J. Vennik in *Advances in Electronics and Electron Physics* (Ed. L. Marton), Academic Press **1977,** Vol. 43, pp. 139–203.

[2] R. Hoogewijs, L. Fiermans, J. Vennik, *Belvac News* **1987,** *3*, 7; *Chem. Phys. Lett.* **1976,** *37*, 87; *Chem. Phys. Lett.* **1976,** *38*, 471; *J. Phys. C.* **1976,** *9*, L103; *J. Electr. Spectr. & Rel. Phen.* **1977,** *11*, 171; *Surf. Sci.* **1977,** *69*, 273.

[3] M. F. Chung, L. H. Jenkins, *Surf. Sci.* **1970,** *22*, 479.

[4] W. A. Coghlan, R. E. Clausing, *A Catalog of Calculated Auger Transitions for the Elements*, ORNL–TM 3567, Oak Ridge National Laboratory, Oak Ridge, TN **1971.**

[5] K. Siegbahn, C. Nordling, A. Fahlman, R. Nordberg, K. Hamrin, J. Hedman, G. Johansson, T. B. Bergmark, S. Karlson, I. Lindgren, B. Lindberg, ESCA, *Atomic, Molecular and Solid Structure Studied by means of Electron Spectroscopy, ESCA*, Almqvist & Wiksells AB, Uppsala **1967.**

[6] K. W. Nebesny, B. L. Maschhoff, N. R. Armstrong, *Anal. Chem.* **1989,** *61*, 469.

[7] M. P. Seah, in *Practical Surface Analysis, Part 1* (Ed: M. Briggs, M. P. Seah), Wiley, New York **1990,** p. 201.

[8] J. A. Schoeffel, A. T. Hubbard, *Anal. Chem.* **1977,** *49*, 2330.

[9] L. E. Davis, N. C. MacDonald, P. W. Palmberg, G. E. Riach, R. E. Weber, *Handbook of Auger Electron Spectroscopy*, Physical Electronics Inc., **1976.**

[10] P. De Volder, R. Hoogewijs, R. De Gryse, L. Fiermans, J. Vennik, *Surf. & Interface Anal.* **1991,** *17*, 363; *Appl. Surf. Sci.* **1993,** *64*, 41.

[11] D. A. Shirley, *Phys. Rev. B.* **1972,** *5*, 4709.

[12] L. B. Leder, J. A. Simpson, *Rev. Sci. Instrum.* **1958,** *29*, 571.

[13] D. A. Huchital, J. D. Rigden, *Appl. Phys. Lett.* **1970,** *16*, 348.

[14] Ph. Staib, *J. Phys. E* **1972,** *5*, 484.

[15] D. A. Huchital, J. D. Rigden, *J. Appl. Phys.* **1972,** *43*, 2291.

[16] Ph. Staib, *J. Phys. E* **1977,** *10*, 914.

[17] D. W. O. Heddle, *J. Phys. E* **1971,** *4*, 589.

[18] See, for instance, P. Grivet, *Electron Optics, 2nd ed (Eng.)*, Pergamon Press, Oxford, **1972.**

[19] G. A. Proca, T. S. Green, *Rev. Sci. Instrum.* **1970,** *41*, 1778.

[20] W. Schmitz, W. Mehlhorn, *J. Phys. E* **1972,** *5*, 64.

[21] D. Roy, J. D. Carette, *Rev. Sci. Instrum.* **1971,** *42*, 1122.

[22] H. Hughes, V. Rojanski, *Phys. Rev.* **1929,** *34*, 284.

[23] S. Aksela, *Rev. Sci. Instrum.* **1972,** *43*, 1350.

[24] S. Aksela, M. Karras, M. Pessa, E. Suoninen, *Rev. Sci. Instrum.* **1970,** *41*, 351.

[25] H. Z. Sar-El, *Rev. Sci. Instrum.* **1967,** *38*, 1210.

[26] H. Z. Sar-El, *Rev. Sci. Instrum.* **1968,** *39*, 533.

[27] H. Z. Sar-El, *Rev. Sci. Instrum.* **1971,** *42*, 1601.

[28] H. Z. Sar-El, *Rev. Sci. Instrum.* **1972,** *43*, 259.

[29] J. S. Risley, *Rev. Sci. Instrum.* **1972,** *43*, 95.

[30] P. W. Palmberg, G. K. Bohn, J. C. Tracy, *Appl. Phys. Lett.* **1969,** *15*, 254.

[31] E. M. Purcell, *Phys. Rev.* **1938,** *54*, 818.

[32] C. E. Kuyatt, J. A. Simpson, *Rev. Sci. Instrum.* **1967,** *38*, 103.

[33] J. Cazaux, *J. Microsc.* **1987,** *145*, 257.

[34] J. Cazaux, *Surf. Interface Anal.* **1989,** *14*, 354.

[35] M. Tholomier, D. Dogmane, E. Vicario, *J. Microsc. Spectrosc. Anal.* **1988,** *13*, 119.

[36] M. M. El Gomati, M. Prutton, B. Lamb, C. G. Tuppen, *Surf. Interface Anal.* **1988,** *11*, 251.

[37] V. V. Zashkvara, M. I. Korsunskii, O. S. Kosmachev, *Soviet Phys.–Tech. Phys. (English Transl.)* **1966,** *11*, 96.

[38] H. Hafner, J. A. Simpson, C. E. Kuyatt, *Rev. Sci. Instrum.* **1968,** *39*, 33.
[39] See, for instance, A. E. Conrady, *Applied Optics and Optical Design*, Dover Publications, New York **1957,** p. 120.
[40] J. L. Gardner, J. A. R. Samson, *J. Electron. Spectrosc. & Relat. Phenom.* **1975,** *6*, 53.
[41] J. M. Watson, US Patent *3 783 280*, 1 Jan. **1974**.
[42] H. G. Nöller, H. D. Polaschegg, H. Schillalies, Proceedings of the 6th International Vacuum Congress **1974**; *Jpn. J. Appl. Phys.*, Suppl. 2, Pt. 1, **1974**.
[43] J. C. Helmer, N. H. Weichert, *Appl. Phys. Lett.* **1968,** *13*, 266.
[44] P. J. Bassett, T. E. Gallon, M. Prutton, *J. Phys. E* **1972,** *5*, 1008.
[45] K. Yates, R. H. West, *Surf. Interface Anal.* **1983,** *5*, 217.
[46] B. Wannberg, A. Skollermo, *J. Electron. Spectrosc. & Relat. Phenom.* **1977,** *10*, 45.
[47] N. C. MacDonald, J. R. Waldrop, *Appl. Phys. Lett.* **1971,** *19*, 315.
[48] J. Cazaux, *Surf. Sci.* **1983,** *125*, 335; M. M. El Gomati, A. P. Janssen, M. Prutton, J. A. Venables, *Surf. Sci.* **1979,** *85*, 309.
[49] M. Prutton, I. R. Barkshire, M. M. El Gomati, J. C. Greenwood, P. G. Kenny, R. H. Robert, *Surf. Interface Anal.* **1992,** *18*, 295.
[50] See, for instance, D. Briggs, M. P. Seah (Eds), *Practical Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, Wiley, New York **1983,** p. 128.
[51] B. L. Chaney, *Surf. Interface Anal.* **1987,** *10*, 36.
[52] K. Sieghbahn, C. Nordling, A. Fahlman, R. Nordberg, K. Hamrin, J. Hedman, G. Johansson, T. B. Bergmark, S. Karlson, I. Lindgren, B. Lindberg, *ESCA, Atomic, Molecular and Solid Structure Studied by Means of Electron Spectroscopy*, Almqvist & Wiksells AB, Uppsala **1967**.
[53] C. J. Hovland, *Appl. Phys. Lett.* **1977,** *30*, 274.
[54] D. J. Keast, K. S. Dowing, *Surf. Interface Anal.* **1981,** *3*, 99.
[55] M. P. Seah, G. C. Smith, *Surf. Interface Anal.* **1988,** *11*, 69.
[56] N. Gurker, M. F. Ebel, H. Ebel, *Surf. Interface Anal.* **1983,** *5*, 13.
[57] N. Gurker, M. F. Ebel, H. Ebel, M. Mantler, H. Hedrich, P. Schön, *Surf. Interface Anal.* **1987,** *10*, 242.
[58] D. A. Allison, T. F. Anater, *J. Electron Spectrosc.* **1987,** *43*, 243.
[59] P. Coxon, J. Krizek, M. Humpherson, I. R. M. Wardell, *J. Electron Spectrosc. & Relat. Phenom.* **1990,** *52*, 821.
[60] E. Adem, R. Champaneria, P. Coxon, *Vacuum* **1990,** *7*, 1695.
[61] G. Beamson, H. Q. Porter, D. W. Turner, *J. Phys. E* **1980,** *13*, 64.
[62] G. Beamson, H. Q. Porter, D. W. Turner, *Nature* **1981,** *290*, 556.
[63] P. Kruit, F. H. Read, *J. Phys. E* **1983,** *16*, 313.
[64] T. Tsubai, E. Y. Xu, Y. K. Bae, K. T. Gillen, *Rev. Sci. Instrum.* **1988,** *59(8)*, 1357.
[65] T. Hsu, J. L. Hirshfield, *Rev. Sci. Instrum.* **1976,** *47*, 236.
[66] N. M. Forsyth, P. Coxon, *Surf. Interface Anal.* **1994,** *21*, 430.
[67] G. C. Hembree, J. A. Venables, *Ultramicroscopy* **1992,** *47*, 109–120.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

2.5 Scanning Microanalysis

2.5.1 Physical Basis of Electron Probe Microanalysis

Classical electron probe microanalysis (EPMA) is carried out by determining the intensity of characteristic X-ray emission from an element of interest in the specimen and comparing it with that from a standard of known composition. Maintaining the same analysis conditions enables an X-ray intensity ratio to be obtained which, according to Castaing's first approximation, relates quantitatively to the mass concentration of the element [1].

2.5.1.1 Electron Interactions in Solids

In general, electron microanalysis is based on the effects of elastic and inelastic scattering of an accelerated electron beam upon interaction with atoms and electrons of the material to be examined [2]. The final signals used for image formation and for analytical measurements are not normally the result of single scattering processes but of some electron diffusion caused by the gradual losses of electron energy and by some lateral spreading due mainly to multiple elastic large-angle scattering. Figure 1 illustrates schematically the most important interaction processes and signals detected in different operating modes of the analytical electron microscope (a) and their information volumes (b). The complete energy spectrum comprises primary electrons with energy E_0, and emitted electrons, ions, heat, quanta, and internally generated signals such as: transmitted electrons ($E \approx E_0$); wide-angle and narrow-angle elastically scattered and/or diffracted electrons ($E \approx E_0$); backscattered electrons ($50\,\text{eV} < E \leqslant E_0$); secondary electrons ($0 < E \leqslant 50\,\text{eV}$); inelastically scattered electrons with energy loss ΔE ($E = E_0 - \Delta E$); sample current or absorbed electrons ($E = E_F$, where E_F is the Fermi level); Auger electrons ($E < 10\,\text{eV}$); hard and soft characteristic and continuous X-ray photons ($0 < h\nu < E_0$); cathodoluminescence ($0 < h\nu < 1\text{–}3\,\text{eV}$); and electron plasmons and lattice phonons.

Usually only a comparatively small fraction of the characteristic X-rays isotropically emitted from the specimen is detected because of small solid angles of collection (10^{-3}–10^{-1} sr). Moreover, inner-shell ionizations result in the emission not only of X-rays but also of Auger electrons, and the X-ray fluorescence yield

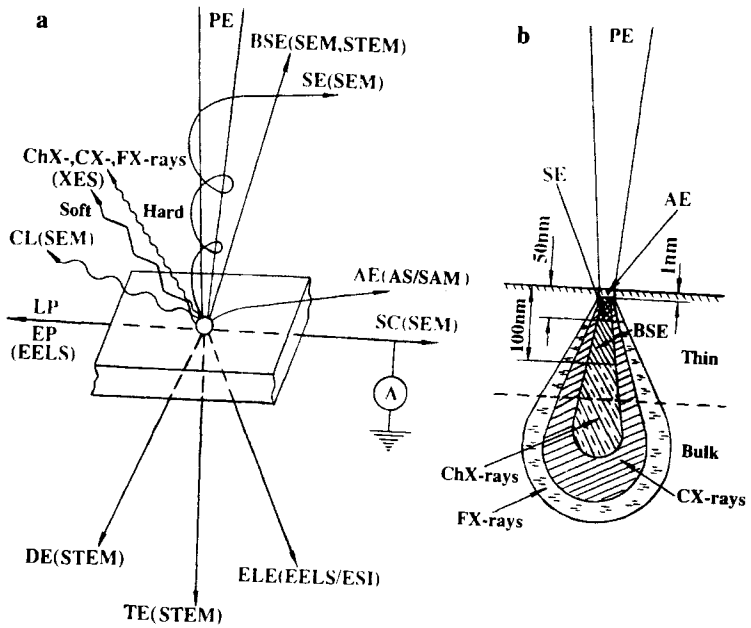


Figure 1. Main processes of electron beam interactions and signals detected in different operating and analytical modes of an analytical electron microscope (a) and schematic arrangement of their information volumes (b); PE, primary electrons; TE, transmitted electrons; DE, diffracted electrons; ELE, electrons with losses of energy; BSE, backscattered electrons; SE, secondary electrons, the information depth of SE is 5–50 nm; AE, Auger electrons, the information depth of AE is about 1 nm; SC, sample current; ChX-rays, characteristic X-rays; CX-rays, continuous X-rays; FX-rays, secondary X-ray fluorescence; CL, cathodoluminescence in ultraviolet, visible, and infrared regions; EP, electron plasmons; LP, lattice phonons. SEM, scanning electron microscopy; STEM, scanning transmission electron microscopy; XES, X-ray emission spectroscopy; AS, Auger spectroscopy; SAM, scanning Auger microscopy; EELS, electron energy-loss spectroscopy; ESI, electron spectroscopic imaging.

decreases with decreasing sample atomic number Z, being below 1% for light elements. A large fraction of about 10–70% of the inner-shell ionization processes leads to inelastically scattered electrons concentrated in small scattering angles which also pass through the objective diaphragm. Atomic electrons can be excited from an inner K, L, or M shell to unoccupied energy states above the Fermi level, resulting in a characteristic edge in the energy-loss spectrum. Complementary to X-ray spectroscopy, in specimens with thickness of about or smaller than the mean free electron path, the well-defined ionization edges, in particular those due to the K shell excitation for elements with atomic number $Z \leqslant 12$, can be easily analyzed by electron energy-loss spectroscopy (EELS) [3].

The plasmons and inelastic intra- and interband excitations of the outer shell electrons near the Fermi level that normally can be observed with energy losses ΔE smaller than 50 eV are influenced by chemical bonds and the electronic band structure, by analogy with optical excitations. In semiconductors the electron impact results in generation of electron–hole pairs and causes an electron-beam-induced current (EBIC). Electron–hole pairs can recombine with emission of

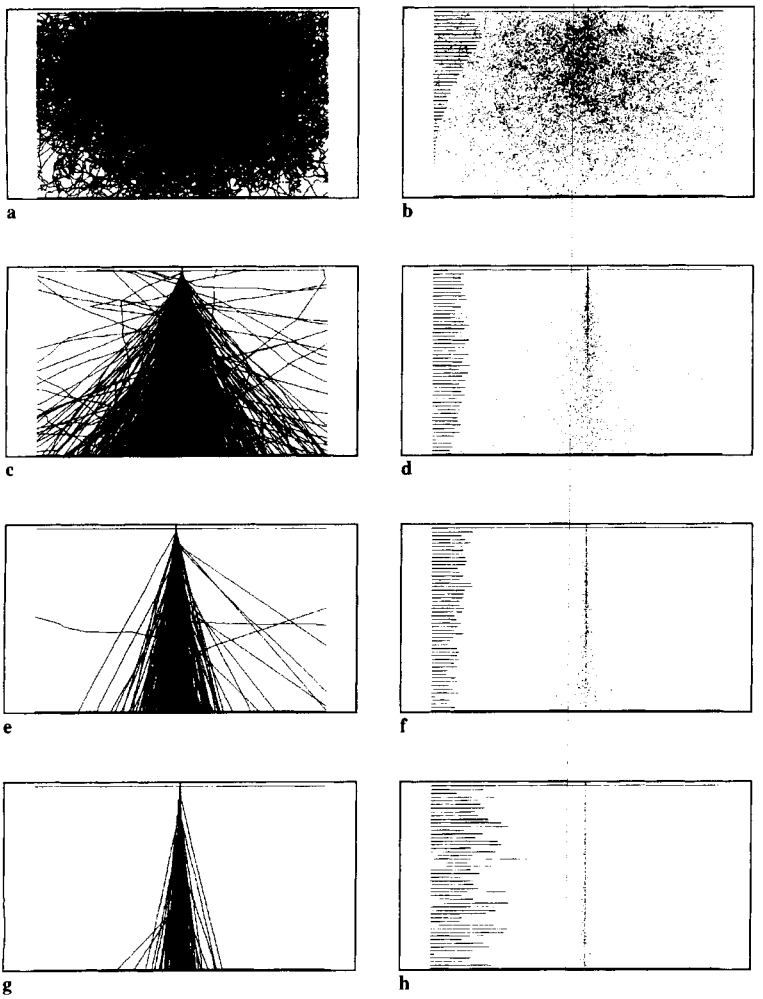


Figure 2. Single scattering Monte Carlo simulations of a lateral electron distribution (a, c, e, g) and Ag L_α X-ray generation (b, d, f, h) in an AgBr tabular crystal of 100-nm thickness with a point electron source at 5 keV (a, b), 30 keV (c, d), 100 keV (e, f), and 300 keV (g, h) for 1000 trajectories. A histogram of the normalized yield of Ag L_α emission $\phi(\rho z)$ is plotted on the left-hand side in (b), (d), (f), and (h), starting at the crystal surface; the length of the bars shows the relative value of $\phi(\rho z)$ at that depth.

luminescent photons in either the ultraviolet, visible, or infrared regions, or by nonradiative lattice phonons. Some fraction of E_0 that is lost in the course of a complicated cascade of inelastic scattering processes is converted into phonons and/or heat and causes radiolysis, thermal damage, bond rupture, and loss of mass and crystallinity by sputtering of specimen matter.

Hence the actual mechanisms of electron–solid interactions may be rather complicated. Therefore, estimates of the information volume of different signals based on Monte Carlo simulations (Fig. 2) are of importance, in particular, for understanding relationships between scattered electrons and X-rays [4]. The consequence of the gradual diminution of the electron energy is that the electrons have a finite depth range of the order of several nanometers up to tens of micrometers, depending on the value of E_0 and the thickness and density of the specimen. The information depth and the lateral extent of the information volume

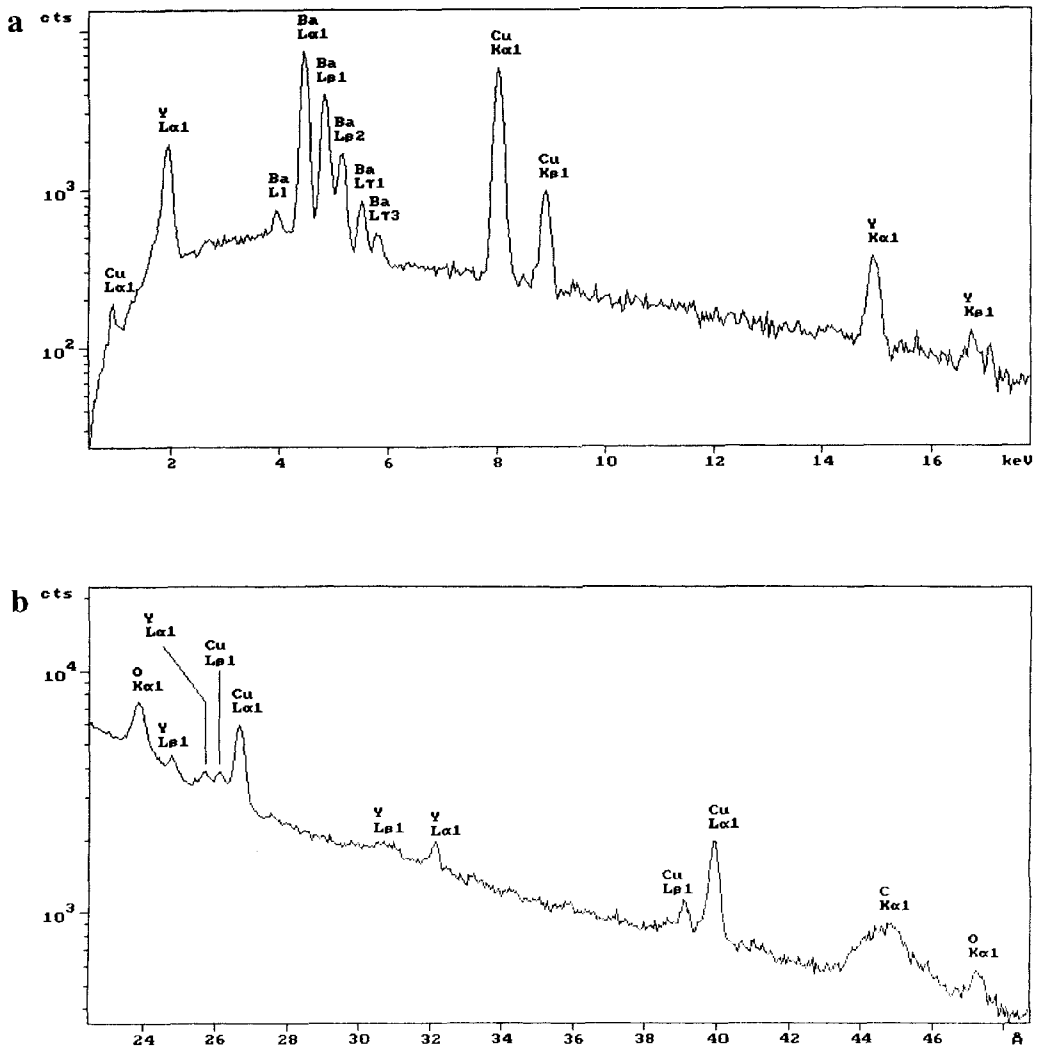


Figure 3. The X-ray emission spectra of a high temperature superconductor $YBa_2Cu_3O_{7-x}$ ceramic recorded at an accelerating voltage of 25 kV. (a) EDS, 1 nA probe current; and (b–d) WDS, 200 nA probe current: (b) lead stearate crystal; (c) PET crystal; (d) LiF crystal.

governing the resolution of the corresponding operating modes contribute to each of the possible signals, decreasing considerably with a decrease in specimen thickness and an increase in its density. In this way the various electron–specimen interactions can generate a great deal of structural and analytical information in the form of emitted electrons and/or photons and internally produced signals.

2.5.1.2 X-Ray Emission Spectra

Ionization of an inner electron shell by the inelastic impact results in a vacancy which

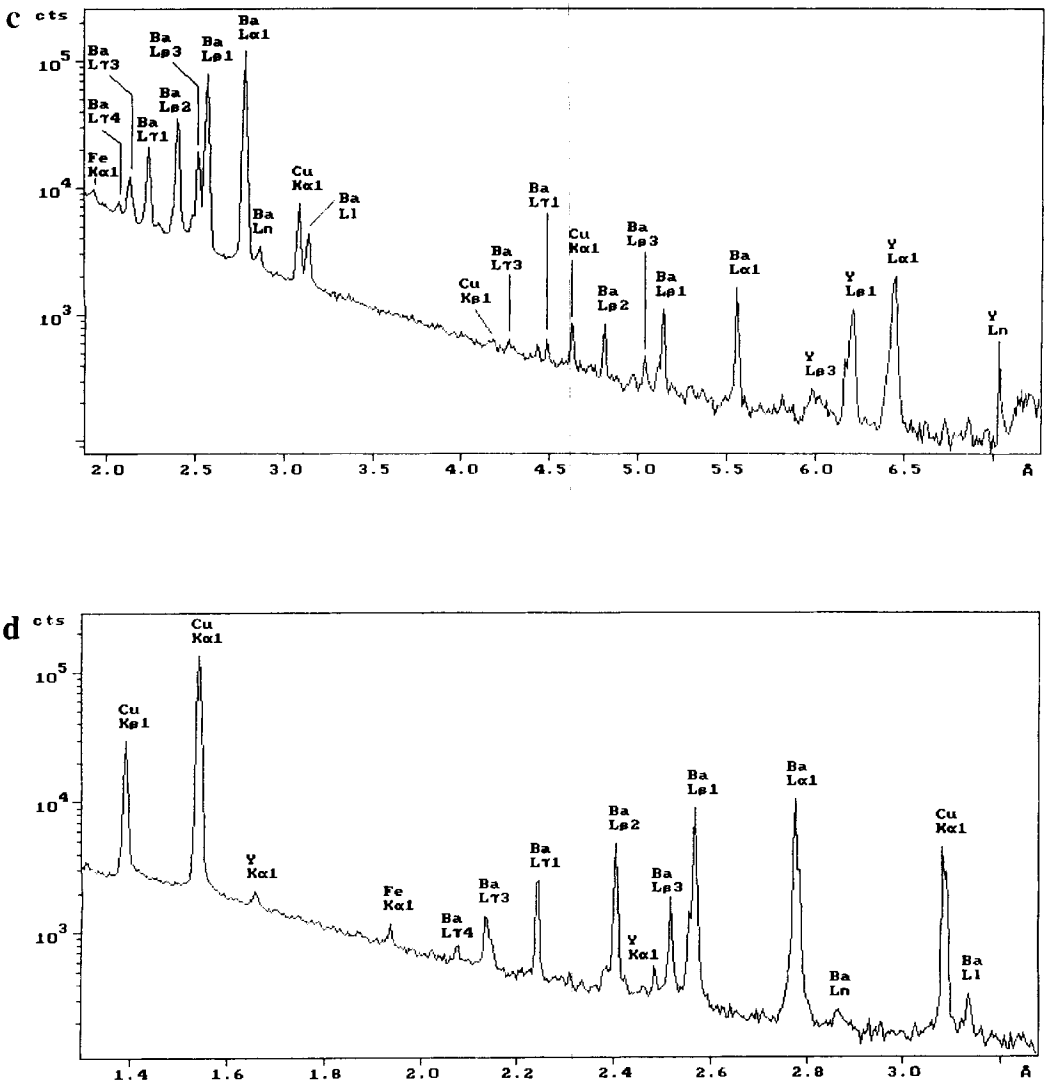


Figure 3 Continued

can be filled by an electron from a higher state. The energy difference can then be emitted either as a characteristic X-ray quantum or as an Auger electron. Hence the spectrum represents the balance between two energy levels; that is, generated intensities are the product of the state density and transition ratio governed mainly by selection rules. The X-ray emission spectra of an advanced high temperature superconductor $YBa_2Cu_3O_{7-x}$ ceramic are shown in Fig. 3. They consist of a background (continuum X-rays, bremsstrahlung) which extends up to the energy of the incident beam, together with superimposed discrete characteristic lines of Cu (CuK and CuL series), Y (YK and YL series), Ba (BaL series) and O (OK

series). The wavelength-dispersive spectroscopy (WDS) spectra (Fig. 3b–d) clearly demonstrate much better resolution than the energy-dispersive spectroscopy (EDS) spectrum (Fig. 3a). The peak of CK_α at 4.47 nm (0.277 keV) in Fig. 3b belongs to a carbon conductive coating deposited on the specimen surface. The X-ray spectra in Figs 3c and 3d also show the presence of traces of Fe (FeK_α at 0.194 nm; 6.398 keV) and Pt (PtL_α at 0.131 nm; 9.441 keV), probably from crucible material.

2.5.1.3 Characteristic X-Ray Spectra

The relationship between the energy of characteristic X-ray emission lines and the atomic number of the element of interest is described by Moseley's law:

$$E = A(Z - 1)^2 \qquad (1)$$

where A is constant within K, L, and M series, and Z is the target atomic number. When an inner shell electron is ejected from the atom, the latter becomes ionized and goes to a higher energy state. The vacancy formed in this way must be filled by an electron from one of the outer levels. Electron transitions are regulated by the selection rule in which $\Delta n \neq 1$, $|\Delta l| = 1, |\Delta j| = 0$, where n, l, and j denote the principal quantum number, azimuthal quantum number, and inner quantum number, respectively. An X-ray quantum can be emitted with a discrete energy corresponding to the difference in energy between the levels involved. Major X-ray emission lines together with their designation are shown in Fig. 4. However, transitions which do not satisfy the selection rules, so called 'forbidden' transitions, can in fact occur and do produce some emission lines, but their intensities are usually low. Bonding also somewhat affects inner shell electrons due to the change in the surrounding charge distribution. However, the range of characteristic lines which allow one to observe chemical effects is fairly limited. This is because the spectrum does not reflect precisely changes in electron states: there are no changes in

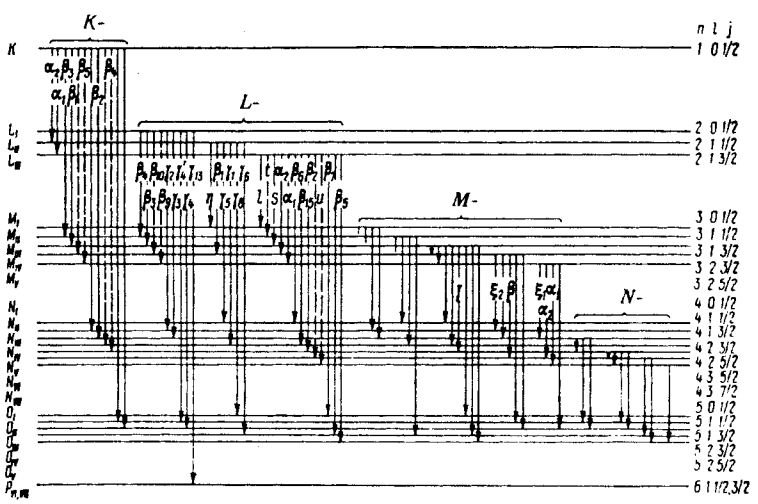


Figure 4. Schematic diagram showing common X-ray emission lines and their designations.

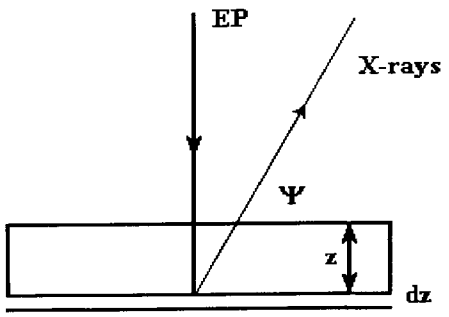


Figure 5. The intensity of the total X-ray emission originating from depth z. EP, electron probe; ψ, angle between emitted X-rays and the specimen surface.

X-ray spectra if the chemical effects on two different energy levels are the same.

The intensity of the total X-ray emission originating from depth z (Fig. 5) below the surface of a specimen with density ρ including any fluorescence contribution may be expressed [1] as

$$I = \phi(\Delta\rho z)\int_0^\infty \phi(\rho z) \times \exp(-\chi\rho z)\,\mathrm{d}\rho z f(\chi)(1+\gamma+\delta) \quad (2)$$

where $\phi(\Delta\rho z)$ is the emission from an isolated thin film of mass thickness $\Delta\rho z$, $\chi = (\mu/\rho)\operatorname{cosec}\psi$, ρz is the specimen mass thickness, and μ is the linear absorption coefficient. The absorption factor is defined as

$$f(\chi) = \frac{\int_0^\infty \phi(\rho z)\exp(-\chi\rho z)\,\mathrm{d}\rho z}{\int_0^\infty \phi(\rho z)\,\mathrm{d}\rho z} \quad (3)$$

The fluorescence correction factor $(1+\gamma+\delta)$ includes the ratio of the fluorescence intensity to the primary characteristic X-ray intensity γ, and the corresponding ratio for the continuum fluorescence contribution δ.

Although most of the characteristic X-ray emission can be explained on the basis of transitions allowed by the selection rules, weak lines may appear which occur as satellites close to one of the principal lines. Their production has been explained by assuming that an atom may be doubly ionized by the incident radiation. The two ionizations have to occur virtually simultaneously because the lifetime of an excited state is very short (10^{-14} s). For example, an electron transition from the doubly ionized state in both K and L orbitals (Fig. 6), results in the emission of a single photon (the KL–LL transition). Its energy may be somewhat higher than that associated with an ordinary L–K transition owing to the fact that the original extra vacancy would have reduced the degree of screening of the nucleus by the electrons and thus increased their binding energy. Satellite lines then appear on the high-energy side of the $K_{\alpha_1\alpha_2}$ peak at energies corresponding to the doubly ionized L suborbital. Two B_2O_3 species in Fig. 6

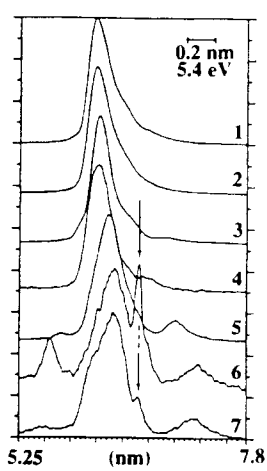


Figure 6. Satellite lines (arrows), line shifts and change of shape due to chemical bonding effect in the BK_α peak of boron containing compounds recorded with a stearate crystal at an accelerating voltage of 10 kV: (1) B in SiC; (2) pure B; (3) LaB_6; (4) BP; (5) BN; (6) B–Si glass (14.9% B_2O_3); (7) datolite (20.24% B_2O_3). Courtesy of JEOL.

(borosilicate glass (6) and natural mineral datolite (7)) reveal the emergence of satellite peaks. However, the satellite intensity in spectrum (7) is much smaller than that in (6), probably due to chemical effects, although the content of B_2O_3 is higher.

Satellites can also occur on the low-energy side of the line and may contribute to the asymmetry of the energy distribution of the band generated. Usually, satellites are relatively more intense for lighter elements because the lifetime of an excited state is longer and the probability of double ionization is higher. Spectral deconvolution accounting for the presence of low- and high-energy satellites and instrumental distortions is required in order to process experimental asymmetrical peaks of soft X-ray emission bands. It has been found that peak shape changes as a function of the excitation conditions, and the matrix composition is related to self-absorption phenomena [5].

2.5.1.4 Soft X-Ray Spectra

The soft X-ray range can be defined as extending from about 100 eV up to 1.5 keV. For light elements ($4 \leqslant Z \leqslant 9$), in particular, there is no alternative other than using soft X-ray emission. However, soft X-rays may be produced not only due to electron transitions involving inner orbitals of the atoms but also as a result of transitions associated with outer orbitals containing valence electrons. The overlap of the valence energy states leads to a decrease in X-ray intensity from ionized atoms. Moreover, because the inner levels are relatively discrete compared with the valence band transitions associated with outer orbitals, the low-energy X-ray lines sensitively reflect the energy states of valence electrons. Chemical bonding effects in the soft X-ray emission spectra are usually more pronounced in insulators than in conductors because the binding energies of valence electrons increase from metallic, through covalent, to ionic bonding. At the same time the energy of the soft X-ray emission decreases correspondingly. Line shifts and change of shape due to the chemical bonding (see Fig. 6) may be observed in the K_α series from the light elements ($Z < 10$), where the L shells involved in K–L transitions are incomplete, as well as in the L series from transition metals and their compounds ($21 < Z < 28$). The BK_α lines of boron in SiC (1), pure boron (2), LaB_6 (3), BP (4), BN (5), and two B_2O_3 containing species (6, 7) show changes in peak position and shape, as well as the above considered satellites. The valence of boron in each compound is +3 and, therefore, it cannot affect the spectra. Furthermore, differences between the BK_α profiles correlate with the electronegativity of the neighboring element P (2.1), N (3.0), and O (3.5) which form the compounds with boron (2.0).

Undoubtedly, analysis of the fine structure of soft X-ray spectra recorded with the appropriate resolution can give not only the elemental composition of an object under study but also important information on electronic structure and bonding. So, analysis of soft $FeL_{III,II}$ X-ray emission spectra of the mineral wüstite allowed an estimate to be made of the relative energies of the valence conduction band orbitals as well as the splitting of these orbitals in the crystal field and the size of their spin splitting [6]. Application

of low-energy X-rays to layered specimens enables one to obtain a reduced depth of ionization. It also permits one to eliminate most of the secondary X-ray fluorescence emission that occurs when lines of higher energy are employed, particularly for the EPMA of multiphase specimens [7]. On the other hand, in this case, quantitative analysis may be complicated by a number of problems such as contamination, coating, background subtraction, and line interference. Therefore, a cautious approach and a deeper insight into the physical processes involved are necessary.

2.5.1.5 X-Ray Continuum

Bremsstrahlung is produced as a consequence of the slowing down of electrons in the Coulomb field of atomic nuclei. The continuous X-rays form a background over a wide energy range $0 \leqslant E \leqslant eU$ extending up to an energy corresponding to the conversion of the entire energy of an incident electron into a radiation photon in one single interaction. The energy distribution of bremsstrahlung may be expressed by the Kramers' equation [1]:

$$N(E)\,\mathrm{d}E = bZ\frac{E_0 - E}{E}\mathrm{d}E \qquad (4)$$

where $N(E)\,\mathrm{d}E$ is the number of photons within the energy interval E to $E + \mathrm{d}E$, $b = 2 \times 10^{-9}$ photons $\mathrm{s}^{-1}\,\mathrm{eV}^{-1}$ electron^{-1} is Kramers' constant, and $E_0 = eU$ is the incident electron energy in electron-volts. The angular distribution of the X-ray continuum is anisotropic.

The bremsstrahlung intensity can be used to calibrate the film thickness in the microanalysis of biological sections. It also contributes to the background below the characteristic X-ray peaks, thereby decreasing the peak-to-background ratio; the latter can be improved by increasing the energy of the incident electron beam.

2.5.1.6 Overview of Methods of Scanning Electron Beam Analysis

EPMA instruments and modern analytical scanning electron microscopes, scanning transmission electron microscopes, and specially designed Auger-electron microanalyzers equipped with corresponding analytical facilities in order to detect X-rays, inelastically scattered electrons, and/or Auger electrons are the most important electron-optical instruments for the analytical and structural investigation of the variety of bulk and thin samples. Table 1 contains some basic data characterizing state-of-art modern analytical scanning electron beam techniques. This information may also be useful in comparing different methods.

2.5.1.7 Electron Probe X-Ray Microanalyzers

The main task of the electron probe X-ray microanalyzer shown in Fig. 7a is to analyze the elemental compositions of flat, polished surfaces at normal electron incidence with a high analytical sensitivity. The ray diagram of such an instrument (Fig. 7b) is similar to that for a scanning electron microscope, but it contains an additional optical microscope to select the specimen points, profiles, and/or

Table 1. Basic characteristics of scanning electron probe methods[a]

Method	Signals	Type of contrast	Resolution (nm) (magnification)	Information obtained	Information depth (nm)	Element range	Detection limit (at. %) (Accuracy (%))	Refs
SEM	SE	Topography, atomic number, orientation	0.6–7 (edge-to-edge) ($\times 10$–10^6)	Structural (surface topography images)	5–50	–	–	2, 8
	BSE	Atomic number 'compo', topography 'topo', orientation	1.5–15 ($\times 10$–10^4)	Elemental (composition images), structural (surface topography images)	100	$\Delta Z < 0.1$	–	
	SC	Atomic number-orientation	3–100	Elemental (composition images), structural	1000	–	–	
STEM	TE, DE TE + DE DE, TE+DE Z contrast	Diffraction Phase	0.2–2 (point-to-point) ($\times 300$–10^7)	Structural (bright-, dark-field images) Structural (lattice images) Structural (dark-field images)				2, 9, 17–19
	BSE	Atomic number	3–10	Elemental	t (film)	$\Delta Z < 0.1$	2–3	
SAED/STEM	DE	–	1000	Structural (point, ring diffraction patterns)	–	–	(10)	2
CBED/STEM	DE	–	5	Structural (microdiffraction patterns)	–	–		2
WDS	X-ray photons	–	1000 (bulk)	Elemental (spot spectra, profiles, mapping)	1000 (bulk)	4–92	0.01 (2–6)	1, 2, 9–12
EDS	X-ray photons	–	1000 (bulk) 0.6–1 (film)	Elemental (spot spectra, profiles, mapping)	1000 (bulk) t (film)	(5)11-92	0.1 (2–6)	1, 9–15

EELS/ESI with $0 < E < 2\,keV$	ELE	–	1–10	Elemental (spot spectra, 2D spectral profiles, mapping); chemical (ELNES, EXELFS) structural (ESI images, diffraction patterns), surface	t (film)	(1)3–92	1–5 (10–20)	9, 11, 16, 17, 19, 20
AES/SAM	SE, AE	–	8–300	Elemental (spot spectra, mapping), chemical (spectra)	1	3–92	0.1 (10)	2, 21
CL, UV-VIS, SEM IR photons $0 < h\nu < 1–3\,eV$	–	–	30–500	Chemical (spectra), structural (spectroscopic images)				2, 24

[a]For abbreviations, see legend to Fig. 1.

(a)

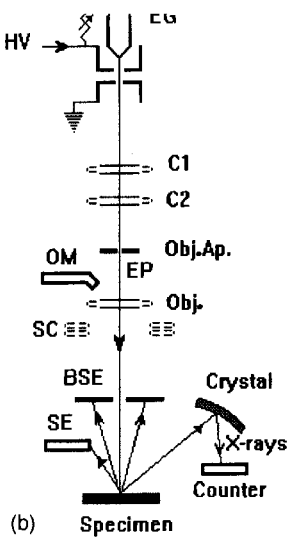


Figure 7. General view (a) and ray diagram (b) of an electron microprobe. On the right-hand side in (b) the electron optical column with two vertical and one inclined wavelength-dispersive spectrometers. On the left-hand side is the data acquisition and processing system including a 32-bit SUN workstation with a 20 in. color screen, a 17 in. frame-store color monitor with a 1024×768 pixel observation screen, and a dedicated keyboard. EG, electron gun; HV, high voltage; C1, first condenser lens; C2, second condenser lens; Obj. Ap., objective lens aperture; OM, optical microscope; EP, electron probe; Obj., objective lens; SC, scanning coils; BSE, backscattered electron detector; SE, secondary electron detector. Courtesy of CAMECA.

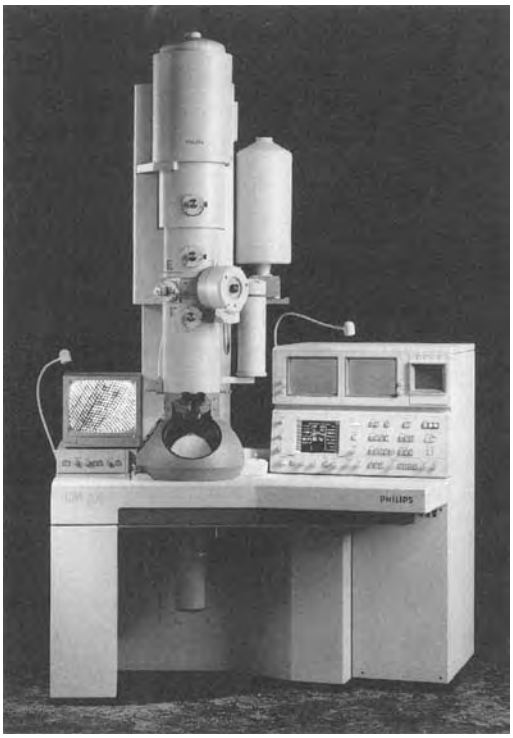
areas to be analyzed and up to five wavelength-dispersive X-ray crystal spectrometers which can record different characteristic X-ray wavelengths, and often also an energy-dispersive X-ray spectrometer which can detect X-rays in a wide energy range. An electron probe (about 6 nm to 1 μm in diameter), governed by the acceptable probe current (10^{-12}–10^{-5} Å), in the scanning mode is produced by a one-, two- or three-stage demagnification of the smallest cross-section of the electron beam after acceleration. Images are displayed on a cathode-ray tube (CRT) rastered in synchronism. The CRT beam intensity may be modulated by any of the different signals (i.e. secondary electrons, backscattered electrons, sample current or X-rays) that result from the electron–specimen interactions.

The combined wavelength- and energy-dispersive microanalyzer is a new-generation analyzer that controls its energy- and wavelength-dispersive spectrometers with the aid of a powerful computer multitasking workstation and presents X-ray data acquired by both spectrometers and images as a unified analysis result. The combination of WDS and EDS can increase the number of simultaneously detectable elements to 13 (5 with WDS and 8 with EDS). A TV display allows the optical microscope image and one of the scanning electron microscope modes (secondary electron or backscattered electron images) to be observed simultaneously.

Traditionally, the energy- and wavelength-dispersive systems in an electron probe X-ray microanalyzer or in an analytical scanning electron microscope are operated at a high voltage (15–40 kV). These conditions are sufficient to allow excitation of X-ray emission of all relevant elements. However, with the increasing interest in detecting light elements and the availability of ultrathin-window or windowless detectors, the importance of the application of low voltages (several kilovolts and less than 1 kV) is also emphasized. With field-emission electron guns that produce sufficient beam currents even in the low-voltage range (5–2 kV), the lateral resolution for X-ray analysis may be improved significantly [22]. As a consequence, thin layers and small particles can be examined in the scanning electron microscope without interference of the bulk support.

2.5.1.8 Analytical Electron Microscopes

A remarkable capability of the scanning transmission electron microscope is the formation of very small electron probes less than 1 nm in diameter by means of a three-stage condenser-lens system. This enables the instrument to operate in the scanning transmission mode with a resolution determined by the electron probe diameter and sample thickness (thin samples). The main advantage of equipping a transmission electron microscope with a scanning transmission electron microscope attachment is the possibility to produce a very small electron probe, with which X-ray analysis can be performed on extremely small areas X-ray generation in thin foils is confined to the small volume excited by the electron probe only slightly broadened by multiple scattering. Better spatial resolution is therefore obtainable for precipitates, or for segregation effects



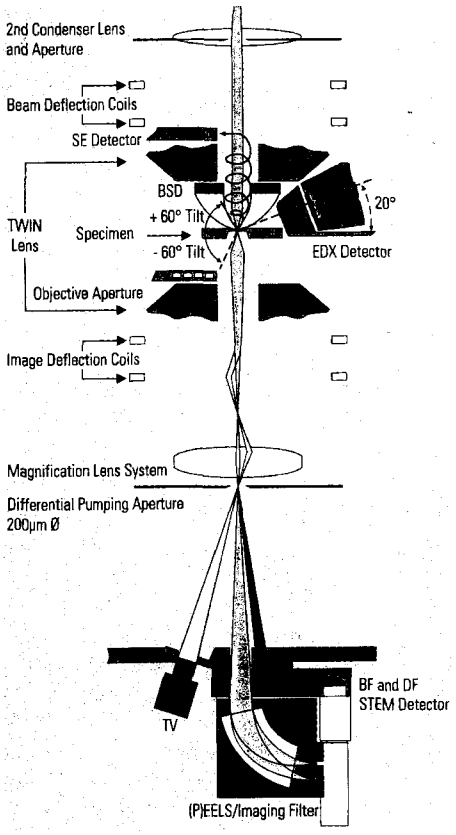


Figure 8. General view (a) and schematic diagram of the electron optics (b) of a dedicated analytical scanning transmission electron microscope. On the left-hand side in (b) is the high-resolution TV monitor and the electron optical column with side-entry airlock and windowless electron detector. On the right-hand side is the scanning observation attachment, control panel and X-ray electron detection system. Courtesy of Philips Electron Optics.

at crystal interfaces, than in an X-ray microanalyzer with bulk specimens, where the spatial resolution is limited to 100–1000 nm by the diameter of the electron interaction volume.

Figure 8 shows a dedicated analytical scanning transmission electron microscope. Normally such an instrument involves a field-emission gun, a probe-forming condenser-lens system, an objective lens, and an electron-detection system, often together with a parallel electron energy-loss spectrometer for separating the currents of unscattered electrons, of elastically scattered electrons and of inelastically scattered electrons. Electron probes of 0.2–0.5 nm diameter (Fig. 9) can be formed, the spherical aberration of the lens being the limiting factor in this case. An advantage of scanning transmission electron instruments is that the contrast can be enhanced by collecting several signals and displaying differences and/or ratios of these by analog or digital processing. Single atoms of heavy elements on an appropriate thin substrate can be

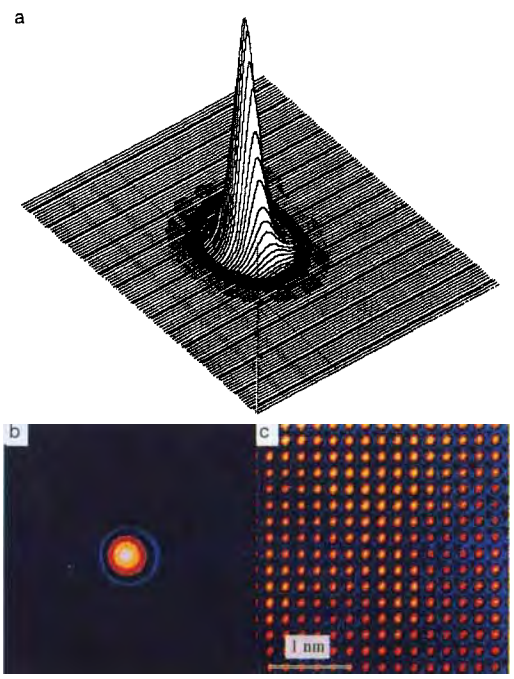


Figure 9. Three-dimensional view of a 0.4 nm (FWHM) electron probe (a) and spot-size measurements at a current of 30 pA (b) using a slow-scan CCD camera. (c) The lattice image of gold in the [100] orientation, recorded and displayed at exactly the same magnification as the image in (b), shows the 0.2-nm spacing for calibration. Courtesy of Dr M. Otten, Philips Electron Optics, The Netherlands.

imaged using Z-contrast with a wide-angle annular dark-field semiconductor detector [19]. A higher contrast than in conventional TEM bright- or dark-field modes is achieved in this case.

X-ray microanalysis in scanning transmission electron microscopy (STEM) has a significant advantage over classical EPMA in that the lateral resolution can be improved by reducing the illumination area to less than 1–10 nm (Fig. 10). The fraction of continuous X-ray emission is lower than for bulk samples because of preferential emission in the forward direction. Recently, improvements in X-ray detection have been achieved by implementing in a single 100 kV field-emission analytical scanning transmission electron microscope several design concepts, including: increasing the collection angle for a solid state detector up to 0.3 sr and the X-ray count rate up to 29 000 cps by the use of electrostatic blanking; and simultaneous X-ray collection from two detectors with equivalent view points and increasing the peak–background value for the intrinsic germanium detector up to 6300 [15]. Determinations of the minimum mass fraction (MMF) for Cr measured on a standard thin film showed that an MMF below 0.1–0.05 wt.% is possible. With dedicated STEM instruments, ultrasensitive analysis of few atoms [13] and at subnanometer lateral resolution has been already realized [14, 19]. Due to the complementary nature of the information obtained, the simultaneous Z-contrast high-resolution imaging and X-ray and energy-loss spectroscopy in the dedicated analytical electron microscope provide a powerful tool for gaining deeper insight into the fundamental correlations between the atomic and electronic structure of materials at the atomic level [20].

2.5.1.9 Multipurpose Electron Probe Analytical Systems

Conventional EPMA utilizes electron excitation for carrying out chemical analysis. However, it is possible to use the electron gun in order to generate an X-ray source for exciting X-rays of sufficient intensity from the specimen; that is, X-ray fluorescence analysis which allows detection limits below 1 ppm to be obtained in an

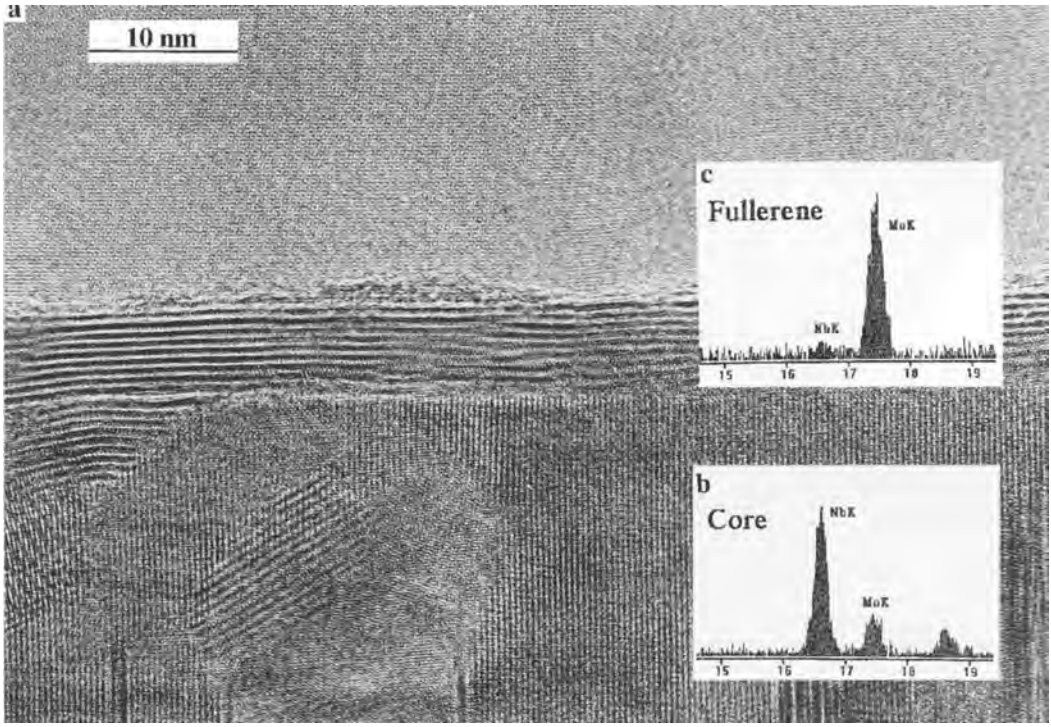


Figure 10. High-resolution image (a) and X-ray energy-dispersive spectra of molybdenum–sulphide fullerene (b) surrounding a niobium–sulphide core (c). Courtesy of Dr M. Otten, Philips Electron Optics, The Netherlands.

instrument supplied with an energy-dispersive spectrometer [23]. A microanalyzer may also be equipped with a cylindrical mirror Auger-electron spectrometer [21]; this, however, needs an ultrahigh vacuum. Auger-electron microanalyzers, in which the 1–10 keV electron gun is incorporated in the inner cylinder of the spectrometer, can work in the scanning mode so that an image of the surface is formed with secondary electrons or an element-distribution map, especially, of light elements is generated using Auger electrons. Digital image processing of multiple detector signals (secondary electrons, backscattered electrons, sample current, Auger electrons, X-rays, cathodoluminescence and/or light image) and computer control with the aid of a powerful multitasking workstation, conversion and storage of data, on-line processing for immediate interpretation of images and spectra, and feedback to the instrument increase significantly the capability of modern analyzers.

Utilizing the scanning beam facilities of the electron probe X-ray microanalyzer or scanning electron microscope, panchromatic cathodoluminescence imaging and cathodoluminescence emission spectroscopy may be easily combined with X-ray energy- and wavelength-dispersive spectroscopies (EDS and WDS), and X-ray mapping and electron imaging to perform comprehensive spatially resolved

microanalysis of point defects in minerals and ceramics and dopant impurities in semiconductors [24]. The environmental scanning electron microscope, the natural extension of a conventional scanning electron instrument that can operate from high vacuum up to a pressure level which can maintain fully wet specimens, is promising for microanalytical studies of uncoated insulating specimens by EDS and cathodoluminescence in conjunction with morphological characterization by secondary electron and backscattered electron imaging [25].

Recently an advanced 300 kV field-emission analytical electron microscope (Argonne National Laboratory) has been designed to attain the best possible analytical sensitivity, resolution, and versatility for EDS, EELS, Auger electron spectroscopy (AES), selected area electron diffraction (SAED), convergent beam electron diffraction (CBED), scanning transmission electron diffraction (ED), and reflected high energy ED consistent for state of the art materials research and still provide moderate imaging capabilities in conventional transmission electron microscopy (high-resolution electron microscopy (HREM)) (CTEM), STEM, and SEM modes. Basically, the system shown in Fig. 11 comprises a conventional field-emission gun with gun lens, a triple condenser, objective, and quadrupole projector. The expected gun brightness is approximately 4×10^9 Å cm^{-2} sr^{-1} at 300 kV and the nominal image resolutions in both TEM and STEM bright-field modes should be better than 0.3 nm (point-to-point), and better than 0.2 nm in the high resolution annular dark-field STEM mode. Figure 12a shows plots of a calculated contrast transfer function for

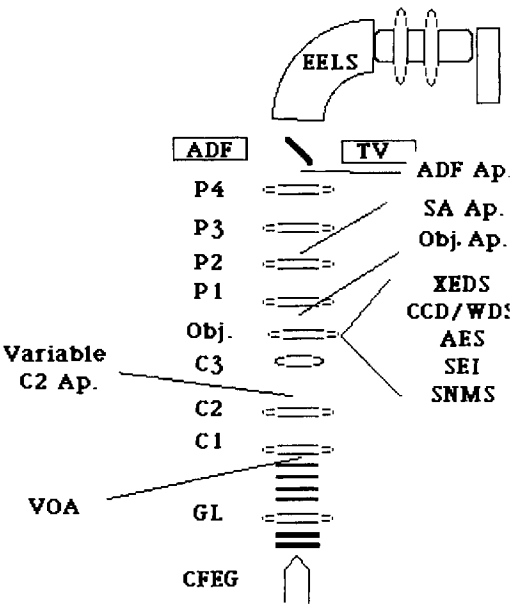


Figure 11. Schematic diagram of the electron optics of an advanced universal analytical electron microscope. CFEG, conventional field-emission gun; GL, gun lens; VOA, virtual objective aperture; C1, C2, C3, condenser lenses; Variable C2 Ap., variable C2 condenser apertures; Obj., objective lens; Obj. Ap., objective aperture; P1, P2, P3, P4, projective lenses; SA Ap., selected-area aperture; ADF, annual dark-field detector; EELS, serial/parallel EELS system; TV, TV video camera; XEDS, windowless X-ray energy-dispersive detector; CCD/WDS, slow scan CCD camera/WDS; AES, conventional hemispherical Auger spectrometer; SEI, secondary electron image detector; SNMS, mini secondary neutral mass spectrometer. Courtesy of Dr N. Zaluzec, Argonne National Laboratory, USA.

the objective lens at 300 and 100 kV, while Fig. 12b shows plots of the calculated probe current/size relationships at 300 kV. The four projectors allow imaging of the probe, which is less than 0.2 nm in diameter, under all conditions.

Specimen rotation about the primary tilt axis is limited to ±85°. Image detection is accomplished by using a video TV camera in the CTEM and bright-field/annular dark-field STEM modes, using one of four operator-selected YAG screens, the signals

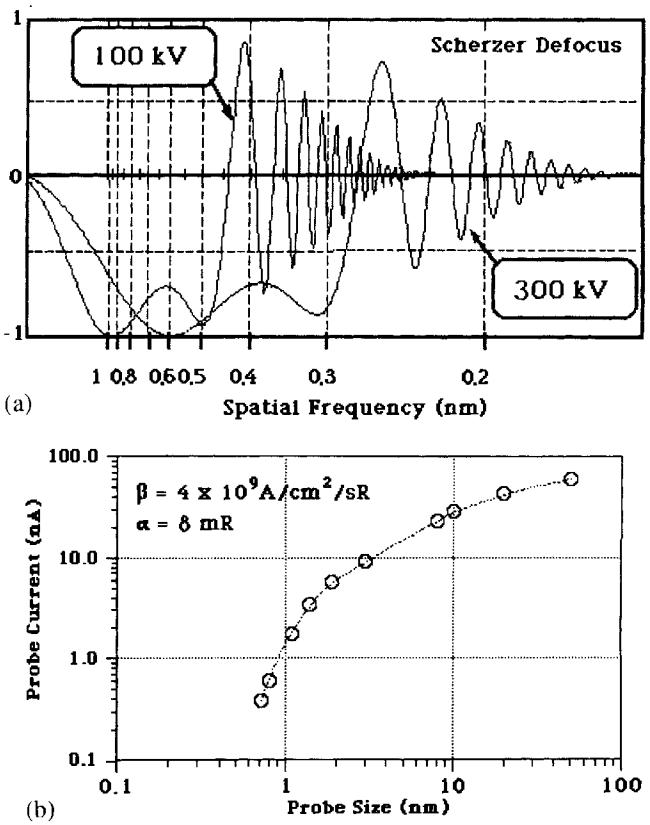


Figure 12. (a) Calculated contrast transfer function at 300 and 100 kV (Scherzer defocus, Cs = 4.4 mm), and (b) probe size/current relationship at 300 kV for the advanced universal Auger electron microscope. Courtesy of Dr N. Zaluzec, Argonne National Laboratory, USA.

from which are flash digitized (8, 16, 32 bit) and routed to two independent frame stores (2K×2K×8 bit). A comprehensive specimen-preparation chamber is interfaced directly to the column; this allows complete extensive cleaning, characterization, and preparation of the specimen surface by the use of sputter cleaning, a mini secondary ion mass spectrometer system, thin film evaporator, etc. Two 400 and two 60 $l\,s^{-1}$ ion getter pumps, four titanium subliminators, and one turbomolecular pump comprises the evacuation system for an ultrahigh vacuum environment (2×10^{-11} to 2×10^{-9} torr).

Multilayer low/high Z material combinations have been employed in both beam- and nonbeam-defining apertures and at all critical surfaces to minimize potential sources of uncollimated hard X-rays which give rise to the hole count phenomenon. The windowless energy-dispersive system has been optimized to maximize the subtending solid angle and allows retraction along a direct line-of-sight path to the specimen. This allows the instrument to achieve a continuously variable solid angle up to a maximum of 0.3 sr. A hemispherical Auger spectrometer with extraction optics is interfaced to the center of the objective lens and both serial and parallel EELS detection capabilities will be present. In addition, secondary and Auger electron spectrometers utilizing parallelizer optics are installed within the objective prefield. This universal AEM

Table 2. Comparison of WDS and EDS [1, 2, 11, 12, 15]

Item	WDS	EDS
Basic method	Wavelength dispersion by diffracting crystals	Energy dispersion by solid state detector
Available elemental range	$_4$Be to $_{92}$U	$(_5B)_{12}$Na to $_{92}$U
Resolution (eV), (MnK_α, FWHM (nm))	~20 (~ 0.7×10^{-3})	100–150 (~6×10^{-3})
Probe current range (Å)	10^{-9}–10^{-5}	10^{-12}–10^{-9}
Detection limit (ppm)	50–100	1500–2000
Number of simultaneously analyzed elements	1	up to 25
X-ray acquisition rate (cps)	10^2–10^5	10^3–3×10^4

system may be controlled either directly by the operator using conventional multifunction dials and switches, or the PC and a mouse directed interactive graphical user interface thus providing telepresence microscopy remote control and operation over networks.

2.5.1.10 X-Ray Emission Spectrometry

Analysis of X-rays emitted from the specimen involves measurements of their energies (or wavelengths) and intensities. There are two main commercially available spectroscopic systems which can be coupled to the scanning electron or scanning transmission electron microscope to record X-ray quanta (i.e. WDS and EDS). Their principal features are compared in Table 2.

2.5.1.11 Wavelength-Dispersive Spectrometry

WDS utilizes the Bragg reflection of X-ray emission dispersed by an analyzing curved crystal from its lattice planes ($2d \sin\theta = n\lambda$). Crystals with different lattice spacing (0.4–10 nm) (Table 3) are used to analyze the whole wavelength range from below 0.1 nm ($UL_\alpha \simeq 0.091$ nm) to above 11 nm ($BeK_\alpha \simeq 11.3$ nm). The spot irradiated by the electron beam on the specimen acts as an entrance slit, while the analyzing crystal and the exit slit are

Table 3. Analysis range of analyzing crystals

Crystal (abbr.)	$2d$ (nm)	Wavelength range (nm)	10 Na	20 Ca	30 Zn	40 Zr	50 Sn	60 Nd	70 Yb	80 Hg	90 Th
STE	10.04	2.22–9.3	$_5$B to $_8$O, $_{16}$S to $_{23}$V								
TAP	2.576	0.569–2.38	$_8$O to $_{15}$P, $_{23}$Cr to $_{41}$Nb $_{46}$Pd---$_{79}$Au								
PET	0.8742	0.193–0.81	$_{13}$Al to $_{25}$Mn, $_{36}$Kr------$_{65}$Tb, $_{70}$Yb---$_{92}$U								
LIF	0.4027	0.0889–0.373	$_{19}$K---$_{37}$Rb $_{48}$Cd----------$_{92}$U								
MYR	8.0	1.77–7.41	$_5$B to $_9$F, $_{17}$Cl to $_{25}$Mn								
LDE1[a]			$_6$C to $_9$F								
LDE2,[a] LDEN			$_5$B to $_8$O								
LDEB[a]			$_4$Be to $_5$B								

[a]Superlattice crystals.

mounted on a Rowland circle. The lattice planes of the crystal are bent so that their radius is $2R$ and the crystal surface is ground to a radius R. Focusing allows one to obtain better separation of narrow characteristic lines and a solid angle of collection of nearly 10^{-3}–10^{-2} sr.

Behind the slit, X-ray photons are recorded by a proportional counter and their energy is converted to a voltage pulse. The detection efficiency of the Bragg reflection and of the proportional counter is about 10–30% [1, 2]. The number of electron–ion pairs generated in the counter is proportional to the quantum energy $E = h\nu$. The pulses are further amplified, discriminated in a single-channel analyzer, and counted by a scaler. The pulse intensity per second is indicated by a ratemeter. Advanced counters with ultrathin mylar windows are available to detect the weak K_α lines of light elements ($4 \leqslant Z \leqslant 11$) more efficiently than in EDS. WDS systems offer much better energy (wavelength) resolution and higher count rates (>50 000 cps) than EDS ones (see Fig. 3 and Table 2). For scanning across a chosen spectral region, the counter and the analyzing crystal should be moved by a pivot mechanism. Usually, X-ray microanalyzers are equipped with several wavelength-dispersive spectrometers which enable different wavelengths to be recorded simultaneously.

2.5.1.12 Energy-Dispersive Spectrometry

With the energy-dispersive spectrometer a solid-state detector is positioned before the dispersing system to collect the distribution of X-ray emission over a wide energy range (0.1–40 keV for a conventional Si(Li) detector, and even up to 80 keV for an intrinsic Ge detector) [10]. Dispersion of the signal detected then takes place by following processing using pulse-height amplification, pile-up rejection of possible coincidence of pulses, and sorting by a multichannel analyzer which relates the measured pulse height to the energy of the incoming photon. The entire assembly of a conventional detector including a field-effect-transistor preamplifier is cryo-cooled under vacuum at near liquid nitrogen temperature to minimize any thermally induced signals. The detection efficiency of an Si(Li) detector is nearly 100% over the range 3–15 keV [1, 2]. The decrease at low energies is caused by the absorption of X-rays in the thin Be window separating the high vacuum microscope column from the detector. Windowless detectors can record K_α quanta from light elements up to $_5$B. At energies higher than 15 keV the decrease in efficiency is caused by the increasing probability of penetrating the sensitive layer of the detector without photoionization.

Recent developments in EDS detector fabrication have resulted in significant characteristic improvements in low-energy performance and sensitivities. Modern EDS detectors can also be exploited without permanent cooling and stay at room temperature without any degradation in performance. Newly designed ones are electrically cooled with an incorporated Peltier device or need no cooling ever, although still at the cost of some decrease in resolution.

Unlike WDS, where the irradiated point has to be adjusted on the Rowland circle, in EDS there is no need for any

mechanical adjustment, and therefore the spectrometer can be used much more effectively for profile or area analyses by scanning large and/or rough specimens. As an energy-dispersive spectrometer occupies a smaller space than a wavelength-dispersive system, it is commonly used in analytical SEM and STEM. A further advantage of EDS is that most of the characteristic X-ray lines can be recorded simultaneously in a short time. For this reason, X-ray microanalyzers that work with WDS are often also equipped with an electron-dispersive spectrometer under the control of the same computer, thus forming an integrated WDS-EDS system in order to provide more rapid and accurate assessments of the elemental composition of the specimen. This includes: simultaneous data acquisition of up to 13 elements with the combination of five wavelength-dispersive and one energy-dispersive spectrometer, more than twice the number measurable with WDS alone; microanalysis of light elements and microvolume elements with a large beam current of 10^{-9}–10^{-5} A (WDS) and of heavy elements of a few percent under the same conditions as for SEM observations at a current of 10^{-12}–10^{-9} A (EDS) for increasing the total analysis efficiency; and preliminary evaluation of beam-sensitive samples with EDS.

2.5.1.13 X-Ray Mapping

X-ray maps (Fig. 13) provide valuable information on the two-dimensional elemental distributions over bulk or thin samples. X-ray mapping can be undertaken with both EDS and WDS, where incoming counts for elements of interest are fed back into the SEM or STEM system. The X-ray analyzer takes control over the electron beam via a special interface, driving the beam around in a rectangular frame and collecting the X-ray emission for each pixel separately. An external beam deflection interface allows X-ray mapping by beam rocking in a transmission electron microscope using standard hardware and software [26].

Color compositional maps provide multielement spatially resolved X-ray analysis within the chosen region of the sample. However, it is often difficult to visualize the compositional ranges and resulting inter-element correlations from the map, particularly for minor or trace constituents. As an alternative method, a composition–composition histogram displaying the numerical relationship between concentrations of the components in various points of the sample can be proposed [12].

The SEM images in secondary and/or backscattered electrons are normally used to choose regions for subsequent X-ray mapping. If features of interest are not differentiated in the monochrome, black and white electron image, they can easily be overlooked for further analysis. Using the X-ray spectrum detected by EDS, one can construct a color response directly related to the underlying elemental composition where the spectrum from each compound has a characteristic color [27]. Thus, the topography and elemental composition of a specimen are compressed into a single view. Furthermore, using this new technique it is not necessary to monitor the X-ray spectrum, set windows, or collect X-ray maps. Compositional data, even for samples with a rough surface, are automatically acquired into the SEM images

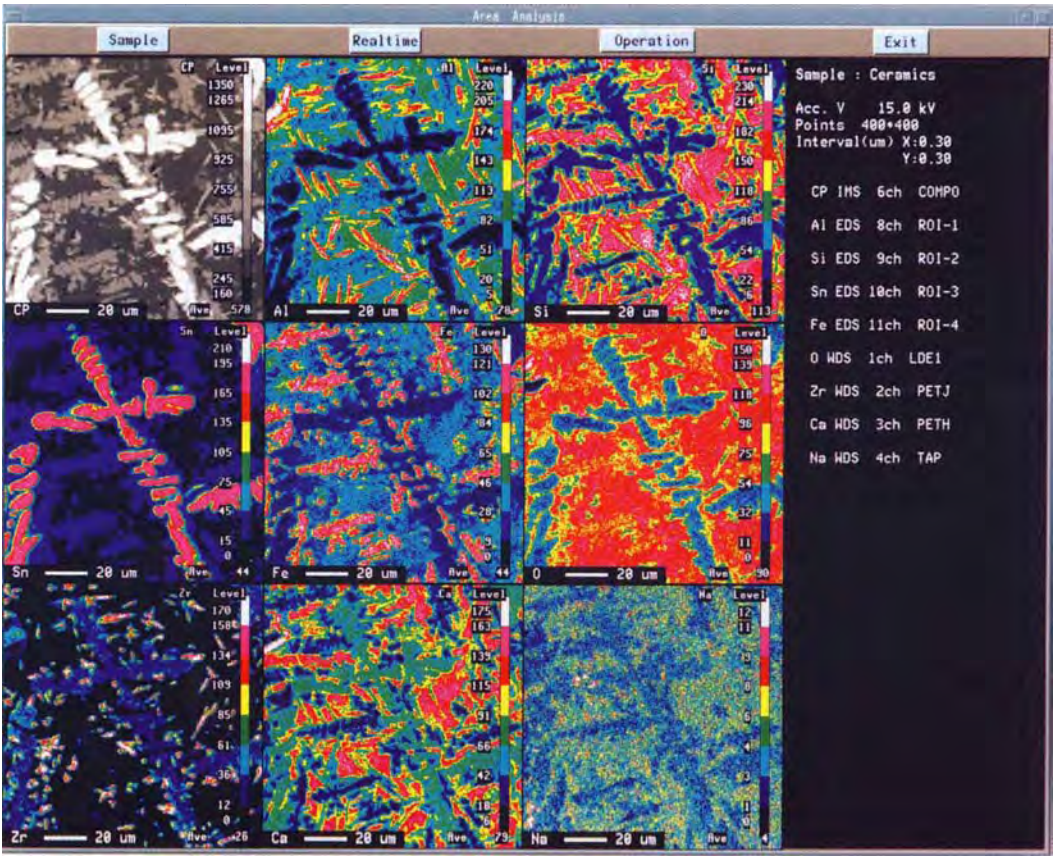


Figure 13. Wavelength- and energy-dispersive combined $140 \times 140\,\mu m^2$ area analysis of an Na–Al–Si–Ca–Fe–Zr–Sn–O ceramic (400×400 points). X-ray maps of eight constituent elements and a backscattered electron image ('compo' mode, CP) recorded at an accelerating voltage of 15 keV. Mapping was performed simultaneously by EDS (Al, Si, Fe, Sn) and WDS (O, Na, Ca, Zr). Courtesy of JEOL.

because the colors displayed always relate to the sample composition.

2.5.2 Introduction to Quantitative X-Ray Scanning Microanalysis

In order to quantify X-ray spectra, the measured intensity of a particular characteristic X-ray line from the specimen should be compared with that from a reference standard of known composition. Hence by keeping instrumental settings (probe current, high voltage, detector efficiency, etc.) constant while the X-ray intensity readings are being taken, one can consider only the ratio of the X-ray intensities of line i of an element a (k_a^i) measured in the specimen I_{sp}^i and the standard I_{st}^i:

$$\frac{I_{sp}^i}{I_{st}^i} = k_a^i \tag{5}$$

where the intensity ratio can be related to the mass concentration c_a of the analyzed element as $k_a^i = f(c_a)$ and $\sum c_a = 1$. It is of course necessary to apply dead-time corrections related to the spectrometer.

An accurate quantitative evaluation can be performed if the k_a^i ratio is corrected for various effects: the atomic number correction for the differences between the electron scattering and penetration in the sample and the standard; the absorption correction for the difference in the absorption of the emitted X-rays as they pass through the sample or standard; and the fluorescence correction for the X-ray fluorescence generated by the X-ray emission in the specimen and in the standard.

2.5.2.1 ZAF Method

Following from Eq. (2), the ratio of X-ray intensities emitted from element a in the specimen and in the pure standard is given by

$$\frac{I_{\mathrm{sp}}^a}{I_{\mathrm{st}}^a} = \frac{\phi(\rho z)_{\mathrm{sp}}^a}{\phi(\rho z)_{\mathrm{st}}^a} \frac{\left[\int_0^\infty \phi(\rho z)\,\mathrm{d}\rho z\right]_{\mathrm{sp}}^a}{\left[\int_0^\infty \phi(\rho z)\,\mathrm{d}\rho z\right]_{\mathrm{st}}^a} \times \frac{f(\chi)_{\mathrm{sp}}^a}{f(\chi)_{\mathrm{st}}^a} \frac{(1+\gamma+\beta)_{\mathrm{sp}}^a}{(1+\gamma+\beta)_{\mathrm{st}}^a} \tag{6}$$

The first term in Eq. (6), which corresponds to the ratio of the emission of element a from an isolated thin layer of mass thickness $\mathrm{d}\rho z$ in the sample and in the standard, respectively, is proportional to the number of ionizations produced by the electron beam:

$$\frac{\phi(\rho z)_{\mathrm{sp}}^a}{\phi(\rho z)_{\mathrm{st}}^a} = \frac{Q\omega(Nc_a/A)\,\mathrm{d}\rho z}{Q\omega(N/A)\,\mathrm{d}\rho z} = c_a \tag{7}$$

where Q is the ionization cross-section of atoms a, ω is the X-ray emission yield, N is Avogadro's number, A is the atomic weight of a, and c_a is the mass concentration of element a in the specimen. Then one can easily deduce from Eq. (6) that the next three terms may be used to represent the atomic number, absorption, and fluorescence (ZAF) correction factors k_z, k_{A}, and k_{F}, respectively. The corresponding corrections are considered in the ZAF method as independent multiplicative terms to the k ratios:

$$k_a^i = k_z k_{\mathrm{A}} k_{\mathrm{F}} c_a \tag{8}$$

2.5.2.2 Atomic Number Correction

The atomic number correction should be applied to the k ratio to compensate for the difference between the electron retardation and electron backscattering in the sample and standard. k_z for element a can be written as:

$$k_z^a = \frac{R_{\mathrm{sp}}^a}{R_{\mathrm{st}}^a} \frac{\int_{E_{\mathrm{c}}}^{E_0} (Q/S)_{\mathrm{sp}}^a\,\mathrm{d}E}{\int_{E_{\mathrm{c}}}^{E_0} (Q/S)_{\mathrm{st}}^a\,\mathrm{d}E} \tag{9}$$

where the terms R and S refer to the electron backscattering and the electron stopping power, respectively, and Q is the ionization cross-section. The integral limits are from the incident electron beam energy E_0 to the critical excitation band E_{c} for the chosen X-ray line of a. The electron stopping power, defined as the energy lost per unit electron path length in material of density ρ, $S = (-1/\rho)\,\mathrm{d}E/\mathrm{d}x$, may be approximated as

follows

$$S = 78\,500 \frac{Z^a}{A^a} \frac{1}{E} \ln \frac{1.166E}{J^a} \tag{10}$$

where $J = [10.04 + 8.25\exp(-Z/11.22)]Z$ [12].

The expressions for the ionization cross-section Q have the general form

$$Q = C \frac{\ln U}{UE_c^2} \tag{11}$$

where C is a constant and $U = \boldsymbol{E}_0/\boldsymbol{E}_c$ is the overvoltage.

The electron backscattering factor $R = (I_t - I_b)/I_t$ defined as the fraction of total generated X-ray intensity excluding loss to backscatter electrons may be calculated using empirical expressions [28]. In a multielement system, the factor R for element i may be derived from the equation

$$R_i = \sum_j C_j R_{ij} \tag{12}$$

where i represents the element being measured and j denotes the elements in the specimen including i; R_{ij} is the backscatter correction for element i in the presence of element j.

Several tabulations [29, 30] and fitted polynomial expressions [31, 32] are available to estimate values of R for various elements as a function of Z and U.

2.5.2.3 X-Ray Absorption Correction

The absorption correction factor $k_A = [f(\chi)^a_{sp}]/[f(\chi)^a_{st}]$ reflects the attenuation of the X-ray intensity measured by the detector. In general, the correction should be considered because the specimen and the standard are not identical. The basic formulation for the absorption term $f(\chi)$ is given by the Philibert–Duncumb–Heinrich equation [33, 34]:

$$\frac{1}{f(\chi)} = \left(1 + \frac{\chi}{\sigma}\right)\left(1 + \frac{h}{1+h}\right)\frac{\chi}{\sigma} \tag{13}$$

where $\sigma = (4.5 \times 10^5)/(E_0^{1.65} - E_c^{1.65})$ is Lenard's constant and $h = 1.2(A/Z^2)$.

The variable h is dependent on the composition amd must be averaged for the various elements present in multielement specimens as follows:

$$h_i = \sum_j c_j h_j$$

In addition, the mass absorption coefficient μ/ρ for the characteristic line of element i must be the weighted sum over all elements:

$$\left(\frac{\mu}{\rho}\right)_i^{sp} = \sum_j c_j \left(\frac{\mu}{\rho}\right)_i^j$$

Calculation of $f(\chi)$ from Eq. (13) is most accurate for values greater than 0.7.

2.5.2.4 Fluorescence Corrections

The characteristic fluorescence correction is necessary, when the energy $\boldsymbol{E}_j$ of the characteristic X-ray line from an element j is greater than the critical excitation energy $\boldsymbol{E}_{c,i}$ for an element i and this difference $\boldsymbol{E}_j - \boldsymbol{E}_{c,i} < 0.5\,\text{keV}$. This may result in excitation of the characteristic emission of element i by the corresponding characteristic line of element j. The basic formulation of the characteristic fluorescence

correction is given by

$$F_i = \frac{\left(1 + \sum_j I^f_{i,j}/I_i\right)_{sp}}{\left(1 + \sum_j I^f_{i,j}/I_i\right)_{st}} \tag{14}$$

where $I^f_{i,j}/I_i$ is the ratio of the characteristic intensity of element i excited by fluorescence to the electron-excited intensity. The total correction should be summed over all the elements in the specimen. In practice, when the standard is either a pure element or there is no significant fluorescence of element i by other elements in the standard, Eq. (14) can be simplified

$$F_i = \left(1 + \frac{\sum_j I^f_{i,j}}{I_i}\right)_{sp} \tag{15}$$

In addition, the calculation of the continuum fluorescence is relatively complicated, involving integration over the range of E_0 to E_c for each element in the sample. When $f(\chi) > 0.95$, $c_i < 0.5$, and $\bar{Z}_{sp} \neq \bar{Z}_{st}$, the correction can be as large as 2–4%, it should be included for highest accuracy [35].

2.5.2.5 $\phi(\rho z)$ Methods

Quantitative microanalysis can be carried out more suitably by treating the atomic number and absorption factors together rather than as separate entities. The correction procedure based on the integration of X-ray distributions, $\phi(\rho z)$ curves, is attractive for the analysis of low-energy X-ray lines, where $f(\chi)$ is much less than 0.7 and the accuracy of the ZAF method is low. However, results of the analysis based on the $\phi(\rho z)$ methods are dependent on the accuracy of description of the experimental $\phi(\rho z)$ curves (especially in the case of systems with unknown X-ray depth distributions) and/or fitting to the corresponding results of Monte Carlo simulation as well. Several researchers have proposed a number of different parametrizations of $\phi(\rho z)$ curves [1, 36–39].

Among the proposed methods the Packwood–Brown model [40] is now widely accepted. This model, which is based on numerous experimental data, utilizes a modified surface-centered Gaussian function to describe the shape of $\phi(\rho z)$ curves. The Gaussian nature of the curve implies some random distribution of electron-scattering events. The surface modification was applied to take into account that the incoming accelerated electrons require some finite distance range before becoming randomized. The general expression for $\phi(\rho z)$, which is drawn in Fig. 14 for the particular case of the AgL and BrL characteristic irradiation generated in a silver halide tabular crystal, is as follows:

$$\phi(\rho z) = \gamma[\exp(-\alpha^2(\rho z)^2)] \times \left[1 - \left(\frac{\gamma - \phi(0)}{\gamma}\right)\exp(-\beta\rho z)\right] \tag{16}$$

The Gaussian function is modified by a transient function to model an increase of the X-ray production away from the near-surface region. The coefficient α relates to the width of the Gaussian function and γ relates to its amplitude. The term β in the transient is related to the slope of the curve in the subsurface region; this means the rate at which the focused electron beam is randomized through scattering in the sample. The intercept $\phi(0)$ is related to the surface ionization potential. The α, β,

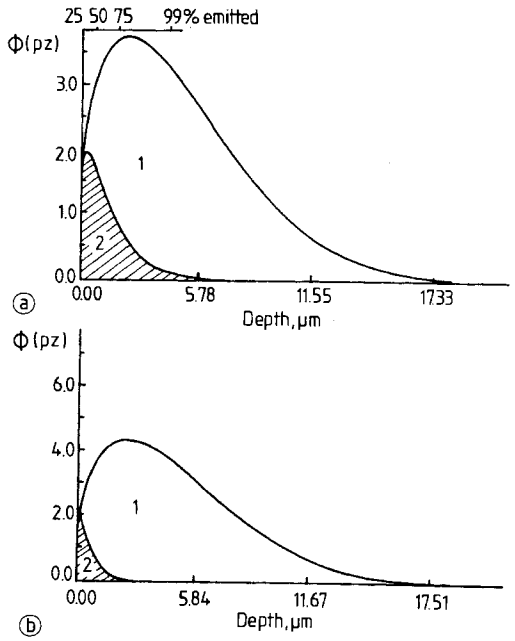


Figure 14. $\phi(\rho z)$ functions for AgL (a) and BrL (b) irradiation produced (1) and emitted (2) under an 80-keV electron beam in an $AgBr_{0.95}I_{0.05}$ tabular crystal; calculated using the PROZA program [41].

and γ parameters are expressed in terms of several different experimental terms (elemental composition of the sample, incident electron energy, atomic number, etc.). Bastin et al. [41, 42] and Armstrong [43] have modified Eq. (16) to optimize the fit for different sets of experimental data.

2.5.2.6 Standardless Analysis

In this approach the measured standard intensities are substituted by calculated ones based on atomic data and empirical adjustments to experimental data. The better quantitative procedures were developed in the last decade in order to validate better standardless calculations. In particular, with the K lines a relative accuracy of 1–5% is reached [44]. Some uncertainties exist with the L lines, and especially with the M lines where atomic data sets are still incomplete or not accurate enough. Further study of the effect of the incompletely filled inner shells is necessary to obtain better atomic data.

The X-ray intensity generated in a bulk sample (standard) for the simpler case of the K line [12, 45] may be expressed as

$$I = g\epsilon p\omega f(\chi)\frac{N}{A}R\int_{E_0}^{E_c}\frac{Q}{\mathrm{d}E/\mathrm{d}\rho s}\mathrm{d}E \qquad (17)$$

where g is a normalization factor which depends on the experimental parameters (beam current, acquisition time, solid angle of the detector, etc.), ϵ is the detector efficiency, p is the transition probability, ω is the fluorescence yield, $f(\chi)$ is the absorption correction defined by Eq. (3), R is the backscatter factor, and Q is the ionization cross-section given in general form by Eq. (12). The additional intensity induced by the secondary X-ray fluorescence is not considered here. When calculating the number of ionizations for the L and M shells, both direct ionizations induced by the bombarding electrons and indirect ones caused by the nonradiative Coster–Kronig transitions between subshells prior to the emission of the X-ray line must also be taken into account. The Si(Li) detector efficiency may be given by

$$\epsilon = \left\{1 - \exp\left[-\left(\frac{\mu}{\rho}\right)_{\mathrm{Si}}\rho d_{\mathrm{Si,act}}\right]\right\} \times \exp\left[-\left(\frac{\mu}{\rho}\right)_{\mathrm{Be}}\rho d_{\mathrm{Be}}\right] \times \exp\left[-\left(\frac{\mu}{\rho}\right)_{\mathrm{Au}}\rho d_{\mathrm{Au}}\right] \times \exp\left[-\left(\frac{\mu}{\rho}\right)_{\mathrm{Si}}\rho d_{\mathrm{Si,dead}}\right] \qquad (18)$$

where Be, Au, Si,act and Si,dead refer to the beryllium window, the gold contact and the active and dead silicon layers of the detector, respectively.

The k ratios obtained by comparing experimental spectra with calculated standard intensities may be used as concentrations of first approximation following the iteration correction [46].

2.5.2.7 Analysis of Thin Films and Particles

Thin films and particles are conveniently analyzed in AEM by using high-energy electron beams with accelerating voltages not less than 80 keV. Under these conditions the effect of electron backscattering is minimal and electron energy losses in the specimen are rather small. For specimens in which the thin-film criterion [47] is fulfilled (i.e. the X-ray absorption <3% and/or fluorescence <5%), the atomic-number correction can be neglected and the characteristic X-ray intensity of element i produced may be expressed [48] as:

$$I_i = K\epsilon_i c_i \omega_i Q_i a_i \frac{t}{A_i} \tag{19}$$

where K is a constant, ϵ_i is the efficiency of the EDS detector to detect X-ray emission from element i, c_i is the mass concentration of element i, ω_i is the fluorescence yield for the analyzed X-ray peak of element i, a_i is the measured fraction of total series intensity, t is the film thickness, and A_i is the atomic weight of element i. To avoid the influence of the substrate, the particles should be deposited onto thin-film-supports of low atomic number such as carbon. Besides, the particles should be scanned in a raster because, in general, the uniformity of the beam density cannot be provided when the probe diameter is comparable to or larger than the particle size.

In practice, it is suitable to measure the relative concentration of element i to another element j which may be given by [49]

$$\frac{c_i}{c_j} = k_{ij}\frac{I_i}{I_j} \tag{20}$$

where c_i and c_j are the mass concentrations of elements i and j, respectively, I_i and I_j are the corresponding X-ray line intensities, and k_{ij} is the Cliff–Lorimer sensitivity factor. The sensitivity factors can be related to an inner standard element as

$$k_{ij} = \frac{k_{i,\mathrm{is}}}{k_{j,\mathrm{is}}} = \frac{k_i}{k_j} \tag{21}$$

where $k_{i(j),\mathrm{is}} = k_{i(j)}$ are the relative sensitivity factors. The index is referred usually to silicon and iron commonly used as internal standard elements. The values of the sensitivity factors for K, L, and M line emission from various elements have been reported by several researchers [50–52].

Equation (20) may be expanded to take into account the effects of absorption and fluorescence:

$$\frac{c_i}{c_j} = k_{ij}\frac{I_i}{I_j}k_{\mathrm{a}}\frac{1}{1 + I_{\mathrm{f}}^i/I_0^i} \tag{22}$$

where I_{f}^i and I_0^i are the fluoresced and primary X-ray intensities of i, respectively, and k_{a} is the absorption factor given by [12]

$$k_{\mathrm{a}} = \frac{(\mu/\rho)_{\mathrm{sp}}^i}{(\mu/\rho)_{\mathrm{sp}}^j} \cdot \frac{1 - \exp[-\chi^i(\rho t)]}{1 - \exp[-\chi^j(\rho t)]} \tag{23}$$

The parameterless extrapolation method proposed by Van Cappellen [53]

is based on measurements of the relative intensities I_i/I_j at various thicknesses following extrapolation to zero thickness ($I_i = 0$) by plotting I_i/I_j versus I_i. Then, if the secondary emission of the elements is much smaller than the primary one, the mass concentration ratio can be determined as follows:

$$\frac{c_i}{c_j} = k_{ij} \lim_{t \to 0} \frac{I_i}{I_j} \tag{24}$$

2.5.3 Conclusions

Since the electron probe X-ray microanalyzer and its basic concepts were first developed in the 1950s by Raymond Castaing, scanning microanalysis using modern X-ray microanalyzers and analytical scanning electron and scanning transmission electron microscopes equipped with energy- and/or wavelength-dispersive spectrometers has been successfully applied to the elemental and compositional characterization of advanced materials in a variety of fields of materials science (metallurgy, ceramics, electronics, geology, etc.). The last decade has shown considerable progress in the theoretical description of fundamental electron–solid interactions and methodological development in scanning microanalysis, especially on fully quantitative and standardless procedures and automatic instrumentation, low-voltage microanalysis of light elements, and super-high sensitive analysis at nano- and subnanometer lateral resolution.

Nowadays, there is a tendency to develop integrated multipurpose AEM systems based on scanning electron and/or scanning transmission electron microscopes equipped with multiple detector systems and to detect effects of inelastic interactions of electron beams within the object (WDS, EDS, EELS and electron spectroscopic imaging, ESI, AS/SAM). However, standard modes, in particular, electron and X-ray diffraction methods (SAED, CBED, electron channelling, Kikuchi patterns), dark field, and stereo-observations are of course still widely used. As a result, conventional SEM, STEM, and EPMA are evolving from instruments for observation or for elemental analysis only, with moderate imaging capabilities, into universal analytical electron microscopy systems for obtaining and treating various types of information on nanometer- and micrometer-sized parts of the object to be examined: its morphology, topography, crystal and defect structure, elemental composition, and electronic state of the matter.

The versatility of advanced electron probe analysis techniques combining high-resolution imaging in various modes, and the powerful analytical facilities and variety of signals and contrast effects available, establish them as outstanding tools for universal applications.

Acknowledgements

The authors are grateful to Dr E. Van Cappellen, Philips Electron Optics, The Netherlands, and to Dr R. Herstens, JEOL (EUROPE) B.V., Belgium, for providing valuable scientific and technical information used in this paper, to Prof. D. Joy, University of Tennessee, USA, for providing the Monte Carlo simulation programs, and to R. Nullens, University of Antwerp (U.I.A.), Belgium, for assistance in recording of X-ray spectra.

2.5.4 References

[1] V. D. Scott, G. Love, S. J. B. Reed, *Quantitative Electron-Probe Microanalysis*, 2nd edn., Ellis Horwood, New York **1995,** 19, 28, 37, 45, 61.

[2] L. Reimer, *Transmission Electron Microscopy. Physics of Image Formation and Microanalysis*, 2nd edn., Springer Verlag, Berlin **1989,** 3, 5, 7, 221, 376, 404, 425, 428.

[3] F. Hofer in *Energy-Filtering Transmission Electron Microscopy* (Ed.: L. Reimer), Springer Verlag, Berlin **1995,** Chap. 4.

[4] D. E. Newbury, *Microbeam Anal.* **1992,** *1*, 9.

[5] G. Remond, C. Gilles, M. Fialin, O. Rouer, R. Marinenko, R. Myklebust, D. Newbury, *EMAS '95 4th European Workshop on Modern Developments in Applied Microbeam Analysis*, EMAS, St Malo, France **1995,** 117.

[6] D. Timotijevic, M. K. Pavicevic, *Mikrochim. Acta* **1992,** *12(Suppl)*, 255.

[7] J. L. Pouchou, *EMAS '95 4th European Workshop on Modern Developments in Applied Microbeam Analysis*, EMAS, St Malo, France **1995,** 95.

[8] D. C. Joy, *Microbeam Anal.* **1992,** *1*, 19.

[9] N. D. Browning, S. J. Pennycook, *Microbeam Anal.* **1993,** *1*, 19.

[10] D. B. Williams, J. I. Goldstein, C. E. Fiori in *Principles of Analytical Electron Microscopy* (Eds.: D. C. Joy, A. D. Romig, J. I. Golgstein), Plenum Press, New York **1986,** Chap. 4.

[11] J. I. Goldstein, D. B. Williams, *Microbeam Anal.* **1992,** *1*, 29.

[12] J. A. Small in *Handbook of X-ray Spectrometry: Methods and Techniques* (Eds.: R. E. Van Grieken, A. A. Markowicz), Marcel Dekker, New York **1993,** Chap. 12.

[13] G. Cliff, R. F. Devenish, P. J. Goodhew, R. J. Keyse, G. W. Lorimer, *Proc. 13th International Congress on Electron Microscopy*, Paris **1994,** Vol. 1, 719.

[14] Y. Bando, *Proc. 13th International Congress on Electron Microscopy*, Paris **1994,** Vol. 1, 591.

[15] C. E. Lyman, J. I. Goldstein, D. B. Williams, D. W. Ackland, S. Von Harrach, A. W. Nicholls, P. J. Statham, *J. Microsc.* **1994,** *176*, 85.

[16] L. Reimer in *Energy-Filtering Transmission Electron Microscopy* (Ed.: L. Reimer), Springer Verlag, Berlin **1995,** Chap. 1.

[17] P. E. Batson, *Proc. 13th International Congress on Electron Microscopy*, Paris, **1994,** Vol. 1, 709.

[18] M. Isaacson, M. Ohtsuki, M. Utlauit in *Introduction to Analytical Electron Microscopy* (Eds.: J. J. Hren, J. I. Goldstein, D. C. Joy), Plenum Press, New York **1979,** Chap. 13.

[19] N. D. Browning, M. F. Chisholm, S. J. Pennycook, *Nature* **1993,** *366*, 143.

[20] S. J. Pennycook, D. E. Jesson, N. D. Browning, M. F. Chisholm, *EMAS '95 4th European Workshop on Modern Developments in Applied Microbeam Analysis*, EMAS, St Malo, France, **1995,** 301.

[21] M. A. Baker, J. E. Castle in *Materials Science and Technology: a Comprehensive Treatment* (Eds.: R. W. Cahn, P. Haasen, E. I. Kramer), VCH, Weinheim **1994,** Vol. 2B, Chap. 13.

[22] C. E. Nockolds, *Microbeam Anal.* **1994,** *3*, 185.

[23] I. Pozsgai, *X-ray Spectrosc.* **1994,** *23*, 32.

[24] J. F. Bresse, G. Remond, B. Akamatsu, *EMAS '95 4th European Workshop on Modern Developments in Applied Microbeam Analysis*, EMAS, St Malo, France **1995,** 213.

[25] G. D. Danilatos, *Mikrochim. Acta* **1994,** *114/115*, 143.

[26] J. M. Brock, *Philips Electron. Optics Bull.* **1994,** *133*, 17.

[27] P. Statham, *EMAS '95 4th European Workshop on Modern Developments in Applied Microbeam Analysis*, EMAS, St Malo, France **1995,** 408.

[28] W. Reuter, *Proc. 6th International Conference on X-ray Optics and Microanalysis*, University of Tokyo Press, Japan **1972,** 121.

[29] G. Springer, *Mikrochim. Acta* **1966,** *3*, 587.

[30] P. Duncumb, S. J. B. Reed in *Quantitative Electron-Probe Microanalysis*, NBS Special Publication No. 298, Department of Commerce, Washington, DC **1968,** 133.

[31] H. Yakowitz, R. L. Myklebust, K. F. J. Heinrich, *NBS Tech. Note 796*, Department of Commerce, Washington, DC **1973.**

[32] R. L. Myklebust, *J. Phys.* **1984,** *45(Suppl 2)*, C2–41.

[33] J. Philibert in *3rd International Congress on X-ray Optics and Microanalysis*, Academic Press, New York **1963,** 379.

[34] J. I. Goldstein, D. E. Newbury, P. Echlin, D. C. Joy, C. E. Fiori, E. Lifshin, *Scanning Electron Microscopy and X-ray Microanalysis*, Plenum Press, New York **1981,** 312.

[35] R. L. Myklebust, C. E. Fiori, K. F. J. Heinrich, *NBS Special Publication No. 298*, Department of Commerce, Washington, DC **1968,** 197.

[36] H. Yakowitz, D. E. Newbury, *SEM I, Proc. 9th Annual SEM Symposium*, IIT Research Institute, Chicago, IL **1976,** 151.

[37] J. L. Pouchou, F. Pichoir, *J. Phys.* **1984,** *45*, C2-17, C2-47.

[38] J. L. Pouchou, F. Pichoir in *Microbeam Analysis — 1988*, San Francisco Press, San Francisco, CA **1988,** 315.

[39] J. L. Pouchou, F. Pichoir in *Electron Probe Quantification* (Eds.: K. F. J. Heinrich, D. E. Newbury), Plenum Press, New York **1991,** 31.
[40] R. H. Packwood, J. D. Brown, *X-ray Spectrom.* **1981,** *10*, 138.
[41] G. F. Bastin, F. J. J. Van Loo, H. J. M. Heijligers, *X-ray Spectrom.* **1984,** *13*, 91.
[42] G. F. Bastin, H. J. M. Heijligers, F. J. J. Van Loo, *Scanning* **1986,** *8*, 45.
[43] J. T. Armstrong, *Microbeam Analysis — 1982* (Ed.: K. F. J. Heinrich), San Francisco Press, San Francisco, CA **1982,** 315.
[44] J. L. Pouchou, *3rd European Workshop on Modern Developments in Applied Microbeam Analysis*, EMAS, Rimini, Italy **1993,** 67.
[45] J. L. Labar, *Microbeam Anal.* **1995,** *4*, 65.
[46] J. L. Pouchou, *Mikrochim. Acta* **1994,** *114/115*, 33.
[47] J. I. Goldstein, J. L. Costley, G. W. Lorimer, S. J. B. Reed in *Scanning Electron Microscopy* (Ed.: I. O. Johari), SEM, Chicago, IL **1977,** 315.
[48] J. P. Goldstein, D. B. Williams, G. Cliff in *Principle of Analytical Electron Microscopy* (Eds.: D. C. Joy, A. D. Romig, J. I. Goldstein), Plenum Press, New York, **1986,** Chap. 5.
[49] G. Cliff, G. W. Lorimer, *J. Microsc.* **1975,** *103*, 203.
[50] J. E. Wood, D. B. Williams, J. I. Goldstein, *J. Microsc.* **1984,** *133*, 255.
[51] T. P. Schreiber, A. M. Wims in *Microbeam Analysis — 1981* (Ed.: R. H. Geiss), San Francisco Press, San Francisco, CA **1981,** 153.
[52] P. J. Sheridan, *J. Electron. Microprobe Technol.* **1989,** *11*, 41.
[53] E. Van Cappellen, *Microsc. Microanal. Microstruct.* **1990,** *1*, 1.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

2.6 Imaging Secondary Ion Mass Spectrometry

2.6.1 Introduction

Secondary ion mass spectrometry (SIMS) is a surface and microanalytical technique for the investigation of the chemical composition of solids. The specimen to be analyzed is brought under high vacuum and bombarded with energetic ions (e.g., 10 KeV Ar^+). The sputtering process results in the emission of secondary ions from the sample surface. These secondary ions are separated according to their mass-to-charge ratio by a mass spectrometer, and the mass selected secondary ions are registered by a suitable detector. SIMS is a surface analysis technique because the secondary ions generated by the sputtering process can only escape from the outermost layers (~2 nm) of the sample.

2.6.1.1 Types of Secondary Ion Mass Spectrometry Measurements

SIMS measurements can be done in various ways. The chemical composition of the sample surface is studied by acquiring a mass spectrum, that is, the intensity of the generated secondary ions is registered as a function of their mass. The sputtering process causes an erosion of the surface. Under prolonged bombardment the surface layer is removed, and deeper layers become exposed to the primary ion beam. By acquiring the intensity of one or more ion species as a function of time and thus as a function of depth, so-called depth profiles are registered. This ability to perform (large-area) surface analysis and depth profiling was originally the main reason for using SIMS as an analytical technique. This functionality proved to be essential in the development of semiconductor materials in the microelectronics industry [1, 2, 3].

Although the ability to map the lateral distribution of secondary ions was already present in one of the earliest commercial available SIMS instruments originally developed by Slodzian [4, 5], the demand to characterize complex structures with micrometer and submicrometer scale dimensions, pushed improvements in instrument design and added imaging capability as a third major functionally to SIMS. By acquiring the intensity of the mass-filtered secondary ions as a function of their location of origin, spatially resolved ion maps can be produced. These ion images reveal the two-dimensional chemical composition of the surface [6]. By combining image acquisition with depth profiling, that is, acquiring images as a function of depth, information on the

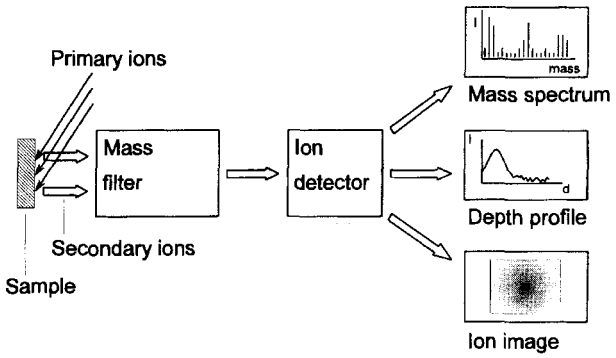


Figure 1. Schematic diagram of a secondary ion mass spectrometer set-up and the basic types of data acquisition: mass scan, depth profile and ion mapping

three-dimensional chemical composition of the sample can be obtained [7, 8]. The essential components of an SIMS instrument and the basic data acquisition modes are shown schematically in Fig. 1.

2.6.1.2 Dynamic and Static Secondary Ion Mass Spectrometry

When dealing with SIMS, one has to distinguish between dynamic and static modes. In a static SIMS experiment the sample receives a total primary ion dose of less than 10^{15} ions cm^{-2}. Under these conditions the erosion of the surface is very limited and only the outermost layer of the substrate is sampled. Static SIMS has a very high surface sensitivity: 0.1% of a monolayer can be detected. It is possible to investigate organic layers: due to the low dose, large molecules can be sputtered from the surface with little or no fragmentation [9].

Dynamic SIMS uses a much higher primary beam current density, up to 0.2 A cm^{-2}, resulting in an erosion rate of several monolayers per second. The high material consumption results in dynamic SIMS having a very high 'bulk' sensitivity. Elements present in the sample at very low concentration levels—parts per million and even parts per billion—can be detected. Depth profiles of typically 1 μm into the sample can be acquired in less than 1 h. However, all information on organic constituents is lost due to the nearly complete fragmentation of the organic molecules under the intense ion bombardment.

2.6.1.3 Ion Microscope and Ion Microprobe

Ion images can be acquired in two different ways: in the ion microscope mode or in the ion microprobe mode [10]. In an ion microscope the sample is bombarded with a broad primary ion beam. The stigmatic ion optical design of the mass spectrometer preserves the lateral distribution of the secondary ions from the point of origin at the surface to the point of detection by a suitable position-sensitive detector. The detector registers a mass-filtered image of the secondary ions as they leave the sample. This is also called direct-imaging SIMS. The lateral resolution of an ion microscope is of the order of 1 μm,

Table 1. Comparison of SIMS with other imaging microanalytical and surface analytical techniques

Technique	Element range	Detection limits	Information depth	Lateral resolution	Type of information
Secondary ion mass spectrometry ion mapping (SIMS)	≥H	ppm	0.3–2 nm	~100 nm	Elemental, isotopic, organic
Scanning Auger microscopy (SAM)	≥Li	0.1%	0.3–2.5 nm	50 nm	Elemental, some chemical states
X-ray mapping in electron probe X-ray microanalysis (EPXMA)	≥B	0.1%	0.5–2 μm	1 μm	Elemental
X-ray photoelectron microscopy (XPS)	≥He	1%	1–3 nm	~30 μm	Elemental, chemical state

and is limited by the aberrations of the stigmatic secondary ion optics.

In an ion microprobe the sample is irradiated by a finely focused primary ion beam, and the mass-filtered secondary ions originating from the location of impact are measured. By scanning the beam over the surface and registering the signal as a function of beam position, an ion image is built up in a sequential manner, much in the same way as in a scanning electron microprobe or a scanning Auger microprobe. The lateral resolution is determined by the diameter of the primary beam, and can be as low as 20 nm.

Both dynamic and static SIMS can be done in the ion microscope and ion microprobe modes. Some instruments are dedicated ion microprobes, whereas others can be operated in both the microscope and in microprobe modes.

2.6.1.4 Characteristics of Secondary Ion Mass Spectrometry

Being a mass spectrometric method, SIMS is able to detect all elements (actually all isotopes) of the periodic table. The sensitivity, that is, the amount of signal generated for a given concentration, varies over orders of magnitude for different elements, and depends strongly on the sample composition and on the experimental conditions. Under favorable conditions detection limits in the part per million and even in the part per billion range are attainable. The high sensitivity of SIMS and the ability to detect all elements, isotopes, and organic molecular fragments are major advances compared to other surface analytical techniques such as Auger spectroscopy and X-ray photoelectron spectroscopy (XPS) [11]. Table 1 compares some characteristics of SIMS with other imaging surface and microanalytical techniques. The ability to map the distribution of chemical species on a microscopic scale with high sensitivity and isotopic discrimination is unique to SIMS [12]. Depth profiling in combination with imaging is a powerful technique to solve problems of multielement characterization of materials at the microscopic level [8], and forms the basis for three-dimensional reconstruction of elemental distributions within sputtered microvolumes of materials [13, 14].

A more detailed treatment of SIMS and its various aspects can be found in monographs [15, 16].

2.6.2 Secondary Ion Formation

When a primary ion with an energy in the kiloelectronvolt range impacts on the surface of a solid it will collide with the sample atoms and gradually lose its energy until it comes to rest a few tenths of nanometers below the sample surface. The energy released in the collision cascade results in a disruption of the material; atoms are displaced and bonds are broken. Particles at and near the surface can receive sufficient kinetic energy to become ejected. Intact molecules, for example adsorbed on the surface, or fragments of these molecules as well as atoms constituting the sample surface or clusters of these atoms, can be ejected. This process is illustrated in Fig. 2. The majority of particles are released as neutral species, but some carry a positive or negative charge. These secondary ions are measured in SIMS.

The sputtering process itself is quite well understood [17]. For the formation of the secondary ion, that is, the ionization process, several models have been put forward, but no general applicable concept is available [18–21]. The observed secondary ion current I_M for an ion M is given by

$$I_M^{\pm} = S\gamma^{\pm} C_M \alpha^{\pm} I_P$$

where S is the sputter yield, $\gamma^{\pm}$ is the ionization probability, C_M is the concentration of species M in the sample, $\alpha^{\pm}$ is the transmission efficiency of the instrument, and I_P is the primary ion current. The $\pm$ sign indicates whether positive or negative secondary ions are considered. The sputter yield S is defined as the total number of sputtered particles per incident primary ion, and depends on the primary beam parameters (mass, energy, and incident angle) as well as on the sample composition. Its value ranges typically between 1 and 10 sputtered particles per incident ion. The ionization probability is the fraction of particles M sputtered as positively (or negatively) charged ions. Its value can vary from ~0.1 to as low as 10^{-5}, and depends on the electronic structure of the species and on the chemical state in which they are before ejection. Electropositive elements (e.g., sodium and the other alkaline elements) have a higher positive ion yield than the noble elements (e.g., gold); the ion yield of silicon is much higher in SiO_2 than in silicon. This complex dependence of the ionization probability on the

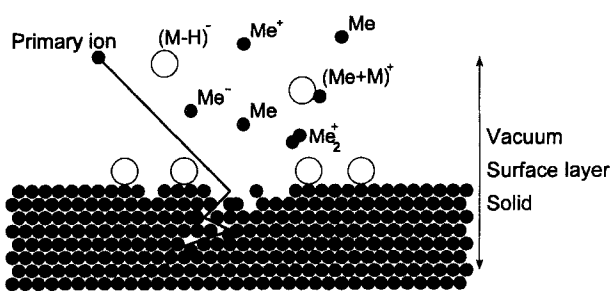


Figure 2. The sputtering process, as a result of the collision cascade of the impacting primary ions, causes, neutral, positive and negative particles to be released from the surface.

experimental conditions and on the chemical state of the sample (the so-called matrix effect) makes truly quantitative SIMS analysis extremely difficult [22, 23]. The ion yield of a species and thus the sensitivity can vary over orders of magnitude depending on the matrix.

The ion bombardment induces considerable changes in the near-surface region of the sample. The impinging ion becomes implanted at a certain depth below the surface. While colliding with the sample atoms, they get displaced and driven deeper into the sample (knock on and mixing), causing a degradation of the surface composition. The process is especially important in dynamic SIMS, and is responsible for the loss of depth resolution during depth profiling.

Whereas the minimal depth resolution of an SIMS measurement is determined by the escape depth of the secondary ions, the ultimate lateral resolution limit is determined by the diameter of the interaction area in the material. Monte Carlo calculations of the intensity of sputtered particles as a function of the radial distance from the impact of Ga^+ ions onto aluminum yield an interaction range of less than 20 nm [24, 25]. A lateral resolution of 20 nm, approaching this theoretical limit, has been obtained with the UC-HRL scanning ion microprobe [26].

2.6.3 Instrumentation

SIMS instruments are rather complex machines costing in the order of US$1 million and having a footprint of some $5\,m^2$ or more. The basic subdivision is between static and dynamic SIMS instruments, although they can also be categorized according to the type of mass spectrometer and image formation.

The four main parts of an SIMS instrument are the primary ion guns, the sample chamber, the mass spectrometer, and the secondary ion detector and the image registration system. Most SIMS instruments use PCs or workstation computers to control all or nearly all instrument functions.

2.6.3.1 Primary Ion Sources

The most important parameters of the primary ion gun are the type of ions, the accelerating voltage, the beam current, and the brightness of the source. Electron impact ion sources, able to produce Ar^+ or O^+ ions, and duoplasmatron sources, able to produce Ar^+, O_2^+, or O^- ions, are low-brightness sources. They can produce beam currents of up to a few hundred milliamps, but the beam can only be focused down to a few micrometres. Cs^+ ions can be produced with a caesium surface ionization source [27]. At ion currents below 100 pA, a submicrometer diameter beam can be obtained. The ultimate beam size and thus lateral resolution in ion microprobes is achieved with Ga^+ liquid metal ion sources [28]. Compared to the chemically neutral Ar^+ ions, O_2^+ primary ions enhance the ion yield of electropositive elements by several orders of magnitude. The implanted oxygen increases the workfunction of the surface, which in turn increases the positive ion formation efficiency. In a similar way, by lowering the workfunction, Cs^+ primary ions enhance the formation of

negative ions from electronegative elements [29].

2.6.3.2 Sample Chamber

To minimize surface contamination during the static SIMS measurements, an ultrahigh vacuum ($<10^{-6}$ Pa) is maintained in the sample chamber. Otherwise, residual gas molecules would deposit on the sample at a faster rate than particles would be removed from the surface by ion bombardment. Due to the continuous erosion of the surface, a dynamic SIMS instrument only requires a moderately high vacuum ($<10^{-4}$ Pa). Sample holders can accommodate specimens of typically up to 1 cm in diameter, although complete wafers can be fitted in some instruments. The sample can be moved in the x, y, and z directions, and rotated. Cooling of the sample at near-liquid nitrogen temperatures is often possible.

2.6.3.3 Mass Spectrometer

Three types of mass spectrometer are in use in SIMS: quadrupole mass filters, double-focusing magnetic sector instruments, and time-of-flight (TOF) mass analyzers [30].

Quadrupole mass spectrometers have a mass range of up to 1000 amu, but only unit mass resolution and a low transmission efficiency (<1%). For imaging they can only be used in ion microprobes.

For the magnetic sector mass spectrometer the mass range is limited to ~500 amu. A mass resolution of up to 10 000 (10% valley) is attainable; however, the transmission efficiency is inversely proportional to the mass resolution, and ranges for ~30% at $M/\Delta M = 800$ to less than 1% at $M/\Delta M = 10\,000$. This mass spectrometer is mainly found in dynamic SIMS instruments and can be used for ion microscopes as well as for ion microprobes.

In a TOF mass spectrometer the mass is determined by measuring the time for which the ions fly in a drift tube after being accelerated by a constant potential. To allow this timing, TOF mass spectrometers are used with pulsed ion sources. This mass analyzer combines a very large mass range of up to 10 000 amu, with a very high transmission efficiency (~50%). The mass resolution is determined by the time uncertainty of the arrival of the ions, that is, the ion formation time (= primary ion pulse width) plus the spread in the initial kinetic energy distribution of the secondary ions. Using subnanosecond primary beam pulses and a drift tube with an electrostatic deflector, a mass resolution of 10 000 (full width at half-maximum, FWHM) at mass 30 is possible [31]. Static SIMS instruments are most often equipped with a TOF mass spectrometer because of the high transmission efficiency. They can be used in ion microprobes and ion microscopes. Table 2 summarizes some of the characteristics of the three mass analyzers.

2.6.3.4 Ion Detection and Image Registration

In an ion microscope the mass resolved and magnified secondary ion image is detected with a position-sensitive device. The first stage of the detection system

Table 2. Characteristics of different mass spectrometers used in SIMS instruments

Mass spectrometer	Mass resolution	Transmission efficiency (%)	Mass range (amu)	Imaging mode
Double focusing and magnetic sector	10 000[a] 800[a]	1 30	0–500	Ion microscope and Ion microprobe
Quadrupole	Unit mass	< 1	0–1000	Ion microprobe
TOF	10 000[b]	~50	0–10 000	Ion microscope and Ion microprobe

[a] 10% valley.
[b] FWHM at $M = 30$.

consists of a single or dual microchannel plate (MCP), which acts as an ion-to-electron converter and amplifier. The gain of the MCP depends on the applied voltage, and is limited to about 1000 for a single MCP and can be up to 10^5 (one ion producing 10^5 electrons) for a dual MCP. The localized electron cloud can be detected with a resistive anode encoder (RAE) or a charge-coupled device (CCD) image sensor or converted to photons using a phosphor screen. In the last case a video camera is used to capture the image and a frame grabber to convert it to a digital image.

The earliest recording of ion images was based on 35 mm film [32]. However, the analog nature of photographic recording does not provide a direct measure of the ion intensities in these images.

In the case of the resistive anode encoder, the electrons from the MCP hit a resistive film (the resistive anode). The position of impact is calculated from the ratio of the charge collected at the four corners of the anode. The impact itself produces a count pulse corresponding to the detection of an individual secondary ion [33]. A schematic diagram of the RAE is shown in Fig. 3.

Video camera-based systems are frequently used for image registration [34–36] as well as digital cameras such as the charge injection device (CID) camera [13]

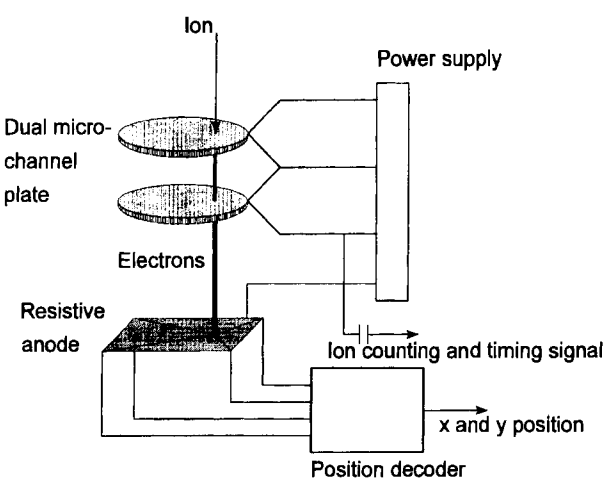


Figure 3. Schematic diagram of an RAE detector used as a position-sensitive ion counter in SIMS imaging.

and the CCD camera [37–39]. The lateral resolution of both the RAE and camera-based system is sufficient not to cause any additional degradation of the ion image.

In an ion microprobe the image registration is in principle much simpler. An ion multiplier is used to count the number of ions of a certain mass while the focused primary beam strikes the sample surface for a certain amount of time. By scanning the beam the image is built up sequentially. Count rates of up to 10^6 counts per second can be sustained. High secondary ion currents need to be measured with a Faraday cage. The dwell time per pixel can be varied from a few microseconds to several milliseconds, resulting in acquisition times between 1 s and tenths of minutes for one image. This is the case when using a quadrupole or magnetic sector mass analyzer. With a TOF mass spectrometer, all secondary ions generated by the impact of one pulse of primary ions are separated in time according to their mass and can be counted one after the other. In this way the images of all masses are built up simultaneously. However, the number of secondary ions per primary ion pulse is very small, zero to a few hundred at the maximum, so that a large number of pulses need to be fired at each location.

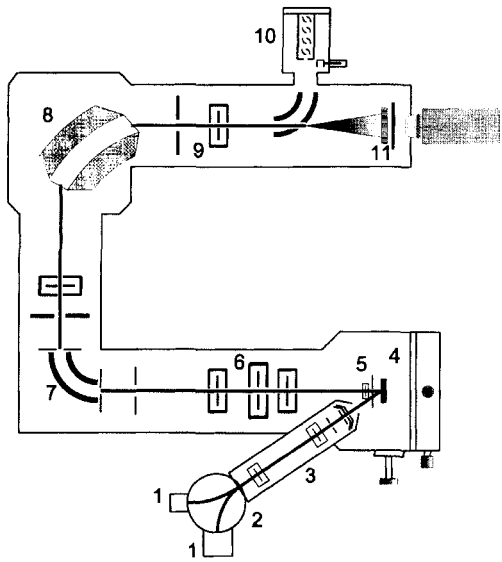


Figure 4. Layout of the Cameca IMS secondary ion mass spectrometer, a dynamic SIMS instrument original designed as an ion microscope but also capable of working as an ion microprobe. 1, ion source; 2, primary beam mass filter; 3, primary column lenses; 4, sample chamber; 5, immersion lens; 6, transfer optics; 7, electrostatic analyzer; 8, magnetic sector; 9, projection lens; 10, electron multiplier and Faraday cage; 11, MCP and fluorescent screen.

2.6.3.5 Typical Configurations

In the following paragraphs we will discuss in some detail two typical SIMS instrument configurations. The Cameca IMS instrument is a dynamic SIMS instrument, originally designed as an ion microscope but also capable of working as an ion microprobe [40, 41]. The Cameca IONTOF SIMS is an example of a static ion microprobe [9, 42, 43].

Figure 4 shows a schematic drawing of the Cameca IMS SIMS instrument. This instrument is a dynamic SIMS instrument for near-surface analysis, depth profiling, and ion imaging. It uses a magnetic sector mass spectrometer of a rather sophisticated ion–optical design that allows the instrument to operate in both the ion microscope and microprobe modes.

In the configuration shown in Fig. 4 the instrument is equipped with two primary ion sources. The duoplasmatron source can be used to produce O_2^+, O^-, or Ar^+ ions. This is a low-brightness source that, for ion imaging, is mainly used in the ion microscope mode. The second source is a

cesium surface ionization source, producing a Cs^+ ion beam with sufficient brightness to allow ion microprobe operation with a lateral resolution of a few hundreds of nanometers. The primary ions are accelerated to an energy that is variable from 5 to 17.5 keV, and pass a magnetic filter which, depending on the magnetic field strength and direction, directs the appropriate ions into the primary column. Three electrostatic lenses focus the primary beam onto the sample. Beam-defining apertures and stigmators in the column determine the shape and current of the beam. At the end of the column, four deflector plates allow scanning of the beam over a sample area of up to $500 \times 500\,\mu m^2$.

The sample is mounted vertically on the sample stage, which can be moved in the x and y directions with the aid of stepping motors. The vacuum in the sample chamber is $\sim 10^{-4}$–10^{-6} Pa. For the measurement of positive secondary ions the sample is kept at a potential of +4.5 kV, and for negative ions at −4.5 kV.

The secondary ions produced at the sample surface are extracted by an immersion lens. The immersion lens front plate is held at ground potential. The transfer optics consist of three lenses, of which only one is energized. This lens focuses the secondary ion beam onto the entrance slit of the mass spectrometer. The mass spectrometer consists of a 90° electrostatic sector and a 90° magnetic sector in a modified Nier–Johnson geometry. By varying the width and position of the energy slit, situated behind the electrostatic sector, secondary ions within a certain narrow energy band (<127 eV) are selected. The spectrometer lens transfers these energy-filtered secondary ions to the magnetic sector. The magnetic field separates the energy-filtered ions into various spatially resolved beams according to the mass-to-charge ratio of the ions. For a given magnetic field strength only secondary ions with a certain mass will pass the exit slit. The width of this slit determines the mass resolution. By varying the magnetic field, ions with different mass can be selected. The mass range is 1–280 amu, or optionally to 500 amu. The mass resolution varies from 200 to a maximum of 10 000 (10% valley).

The projection lenses, situated behind the exit slit of the mass spectrometer, are used to postaccelerate the energy- and mass-filtered secondary ions and to direct them into the detection system. The beam can be focused and deflected into an electron multiplier or a Faraday cage to measure the ion current of the secondary ions. Alternatively, the beam can be projected onto an MCP–fluorescence screen assembly.

In the ion microscope mode, a primary ion beam with a diameter of up to 200 μm and a current density of up to $50\,mA\,cm^{-2}$ is rastered over the sample surface. The primary reason for scanning the beam is to get an even 'illumination' (and erosion) of the sample surface. The immersion lens forms a global (nonmass filtered) image of part of the bombarded surface area. The size of the imaged field (the area from which ions can enter the mass spectrometer) is determined by the transfer lens and the field aperture. Three different transfer lenses and their associated apertures allow image fields with diameters of 25–400 μm, corresponding to a magnification between 2000× and 50×, respectively. The cross-over of the global image, as projected by the transfer optics, is situated

at the entrance slit of the mass spectrometer. Due to lens aberrations, not all secondary ions are focused at the cross-over at one point but rather within a disc of finite diameter. The contrast aperture, placed in front of the cross-over point, is used to reduce the aberration. A smaller aperture results in a higher lateral resolution at the expense of intensity. The global image itself is formed at a fixed distance beyond the slit. The electrostatic sector forms a second virtual global image. The spectrometer lens transfers this to a third virtual image inside the magnetic sector. The energy- and mass-filtered image inside the magnetic sector is transferred by the projection lenses onto the MCP and fluorescent screen.

The process of image formation in the ion microprobe mode is considerably simpler. In this mode the cesium primary ion source is often used to produce a primary ion beam with a diameter of a few hundred nanometers at a current of up to 100 pA. This focused primary ion beam is scanned over the sample area, and the secondary ions released at the point of impact of the primary beam are mass analyzed by the mass spectrometer and measured (counted) by the electron multiplier. For optimal performance in the ion microprobe mode the mass spectrometer must accept a very large fraction of the secondary ions. Since the system must only measure the arrival of the secondary ions and not their position, all apertures and slits are opened.

The Cameca TOF-SIMS instrument is typical of a static SIMS instrument equipped with a TOF mass spectrometer. The schematic diagram of this instrument is given in Fig. 5. One distinguishes again two ion sources, the sample chamber and the TOF mass spectrometer with its read-out electronics.

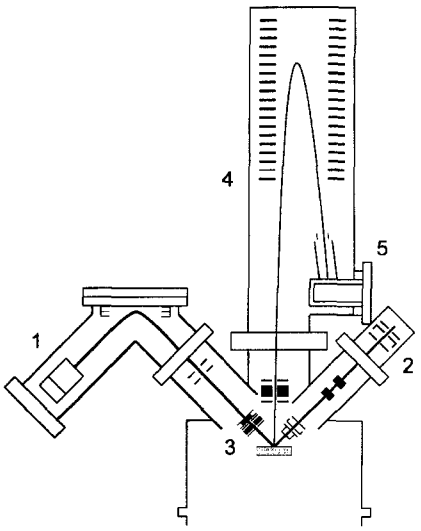


Figure 5. Typical layout of the Cameca TOF-SIMS, a static ion microprobe instrument. 1, electron impact source; 2, gallium liquid metal ion source; 3, sample chamber; 4, flight tube with reflectron; 5, ion detector.

Pulsed ion sources, delivering very short bunches of ions at a high repetition rate, are used. Each bunch of primary ions creates a few secondary ions which are extracted into the TOF mass spectrometer. Their arrival time at the detector is a measure of their mass.

The electron impact ion source can deliver Ar^+ and O^+ ions with an energy of 11 keV. The 90° deflector is used as an electrodynamic mass filter and beam chopper. It also provides axial compression of the ion bunches. An Einzel lens is used for beam focusing. Beam alignment and scanning is done with the X and Y deflectors. The objective lens transfers the beam to the sample surface. Space charge effects impose limits to the beam current (the number of primary ions in each pulse), the pulse length, and the beam diameter.

If, for a given number of ions, the diameter is decreased, the length of the pulse increases. For a beam diameter of 25 μm the minimum pulse length is 0.8 ns, corresponding to ~1800 ions per pulse or a current of 3 pA. At this pulse length the TOF mass spectrometer has a mass resolution of $M/M\Delta$ of about 7000 at mass 29. If the beam diameter is decreased to 3 μm, the pulse length increases to 50 ns, resulting in a mass resolution of ~300, while each bunch contains about 300 ions. The maximum repetition frequency of this source is 10 kHz.

The gallium liquid metal ion source is used in combination with a two-lens column containing a stigmator and a raster unit. The accelerating voltage is 25 kV and the pulse frequency 40 MHz. This ion gun uses an axial electrostatic buncher to produce very short pulses. When the lenses are operated at intermediate cross-over, high lateral resolution can be achieved at high mass resolution. With a beam diameter of ~150 nm and a pulse length of 1.5 ns (four ions or 500 pA) the mass resolution is ~5000. A higher beam current can be achieved with the same spot size by increasing the pulse length (100 ns = 266 ions) but at the expense of mass resolution. The highest lateral resolution is achieved with the beam collimated down to a diameter of 80 nm. With a pulse width of 100 ns (= 30 ions), unit mass resolution is achieved of up to 250 amu.

These pulsed ion sources produce a current density too low to erode the sample (static SIMS conditions). For depth profiling a third wide-beam ion source must be installed. Depth profiling is done in the dual-beam mode, alternating between analyzing the surface with the pulsed source and sputtering the sample with, for example, a 50 μm diameter Ar^+ ion gun.

The ultra-high vacuum chamber has a base pressure of $\sim 6 \times 10^{-8}$ Pa maintained by a turbomolecular pump to ensure an oil-free vacuum. A specimen manipulator can translate the sample in the x, y, and z directions. The sample is at ground potential.

An Einzel lens is used to extract the secondary ions in combination with a quadrupole for secondary beam alignment. The extraction potential is ±2–3 kV.

The TOF mass spectrometer is of the 'reflectron' type with first-order energy focusing. The secondary ions enter the electrostatic reflector, where they are first deaccelerated and then accelerated again, making a nearly 90° turn. The energy of the secondary ions is the sum of their initial kinetic energy and the acceleration they receive when they were formed at or near the surface. Ions with an excess initial kinetic energy will penetrate deeper into the reflector and therefore travel in total a longer distance in the flight tube and arrive at nearly the same time at the detector as ions with the same mass but lower initial kinetic energy. A mass resolution in excess of 10 000 ($M/\Delta M$, FWHM) can be achieved in this way. The mass range extends up to 10 000 amu. Depending on the initial kinetic energy distribution of the second ions the transmission varies between 20 and 80%.

The detection system comprises a single MCP followed by a scintillator and a photomultiplier. The output of the photomultiplier goes to a multistop time-to-digital converter with a time resolution that is variable from 50 to 800 ps, which can detect up to 512 events (= secondary ions) per primary ion pulse. Up to 4000

elemental images can be recorded in parallel.

Two other important configurations are the University of Chicago scanning ion microprobe and the TOF SIMS instrument developed by Charles Evans and Associates.

The University of Chicago UC-HRL scanning ion microprobe is a dynamic SIMS instrument currently having the highest lateral resolution. It uses a Ga^+ liquid metal ion source mounted perpendicularly above the sample surface. Using an acceleration voltage of 40–60 kV, beam diameters of 20–85 nm carrying a current of 1.6–60 pA are obtained. Originally the secondary ions were analyzed with a quadrupole mass spectrometer [44]. Because of the low transmission of the quadrupole mass spectrometer (<0.1%), the instrument has been fitted with a magnetic sector mass spectrometer giving a transmission of ~20% at a mass resolution $M/\Delta M = 1000$ [45].

The TOF secondary ion mass spectrometer developed by Charles Evans and Associates is a static SIMS instrument using a TRIFT mass spectrometer of unique design that allows both ion microscopy and ion microprobe operation in combination with a TOF analyzer [46]. It can be equipped with a large-diameter Cs^+ ion source for ion microscope analysis and with a Ga^+ liquid metal ion source for imaging in the microprobe mode. The TRIFT mass spectrometer has stigmatic optics consisting of three electrostatic energy analyzers [47]. In the ion microscope mode, a mass resolution >5000 can be obtained, while the lateral resolution is limited to ~3 μm. In the ion microprobe mode a submicrometer lateral resolution is attainable at the cost of mass resolution. A time-to-digital converter and a resistive anode encoder are used as an ion counting/position-sensitive detector for the image registration. The instrument operates over a mass range from 1 to 10 000 amu.

2.6.4 Comparison of Ion Microprobe and Ion Microscope Mode

Table 3 compares some of the features of ion microprobes and ion microscopes. As already stated, the lateral resolution of the ion microscope is limited to about 1 μm by the aberrations in the stigmatic ion optics. In the ion microprobe the lateral resolution is determined by the diameter of the finely focused primary ion beam that scans over the sample surface. With

Table 3. Comparison of the main features of ion microprobes and ion microscopes

Feature	Ion microprobe	Ion microscope
Primary ion current	0.1–100 pA	1 nA–1 μA
Beam diameter	<1 μm	100 μm
Analyzed area	10 μm	150 μm
Lateral resolution	~100 nm	~1 μm
Detection limits	~0.1%	~1 ppm

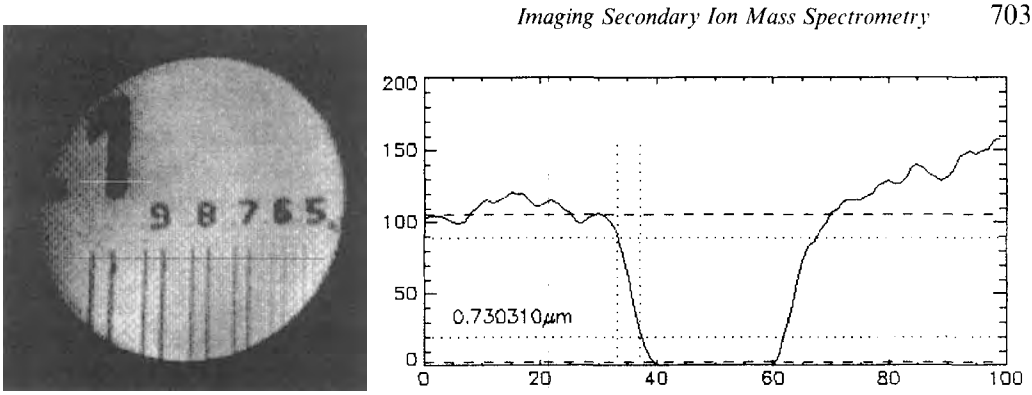


Figure 6. (a) Si^+ ion image obtained from a silicon substrate with a gold test pattern (Cameca 4f in ion microscopy mode) and (b) line scan over the digit '1'. The image has a lateral resolution of 0.7 μm close to the maximum attainable resolution in the ion microscope mode.

high-brightness Ga^+ liquid metal ion sources, a diameter and thus a lateral resolution of ~100 nm is readily obtainable. Figure 6a shows the Si^+ image obtained from a silicon substrate covered with a gold test pattern using a Cameca 4f instrument. The bars have a width from 1 μm down to 0.5 μm. To measure the lateral resolution [48], a line scan over the digit '1' was made (Fig. 6b). The intensity drops from 84 to 16% of the maximum over a distance of 0.7 μm, which is close to the maximum attainable resolution. Figure 7 shows the ion images of Br^- from a silver halide microcrystal acquired with a TOF ion microprobe instrument using a Ga^+ liquid metal ion source with a pulse width of 100 ns at a mass resolution of 300. The image field is $6 \times 6\,\mu m^2$, and the lateral resolution is about 75 nm, clearly demonstrating the superior lateral resolution of the ion microprobe.

Other factors to be considered when comparing the ion microscope and the ion microprobe are mass resolution, analyzed area (or volume), and image acquisition time.

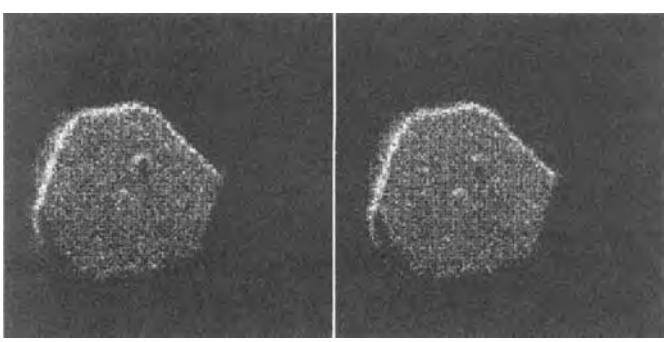


Figure 7. Br^- secondary ion images from a silver halide microcrystal obtained with a TOF SIMS ion microprobe using a Ga^+ liquid metal ion source. The imaged field is 6×6 μm and the lateral resolution is ~75 nm. Because of the high lateral resolution of the ion microprobe, structures at the surface of this micrometer-sized object can be observed.

In the ion microscope the mass resolution is in principle independent of the lateral resolution, although high mass resolution is only obtained at the expense of instrument transmission and thus sensitivity. In ion microprobes equipped with a TOF mass spectrometer, the highest lateral resolution is only obtained at low mass resolution.

The main difference between the ion microprobe and the ion microscope modes of imaging is in the applied primary ion current. The total number of atoms consumed in an SIMS experiment in a time t is given by $N_c \sim A\Phi_P St$, where A is the analyzed area, Φ_P is the primary current density, and S is the sputter rate [49]. The current densities in the ion microprobe and in the ion microscope are of the same order of magnitude, but in the microscope we can use a much broader beam and thus a much higher current. We can analyze an area of $100 \times 100\,\mu m^2$ in the microscope mode using a beam that covers the entire area, and obtain secondary ion images quasi-instantaneously. To analyze the same area in the ion microprobe mode we use a beam of $1\,\mu m^2$ and raster this beam over the surface. Since the current density is about the same in both experiments, we would need to measure 10 000 times longer to sputter the same amount of material (i.e., to obtain the same detection limit). Or, when using the same acquisition time, the ion microscope is orders of magnitude more sensitive than the ion microprobe. The ion microscope is therefore most suitable to analyze large areas with high sensitivity, but if submicrometer lateral resolution is required, the microprobe is the preferred instrument.

2.6.5 Ion Image Acquisition and Processing

2.6.5.1 Dynamic Range of Secondary Ion Mass Spectrometry Ion Images

Image acquisition is complicated by the dynamic range of SIMS, especially in ion microscopy. Secondary ion intensities can easily vary over 3–6 orders of magnitude. To observe high concentrations next to very low ones, an equally large dynamic range is required within one image. Due to both the large variation in concentration and sensitivity for different species, one also has to accommodate the large signal intensity variation from one image (or mass) to another.

The RAE has a dynamic range of about 50 000, which is sufficient for most practical applications. Moreover, this device operates in the ion-counting mode, so that the read-out is a direct measure of the true ion intensity [50].

The situation is quite different when using video camera-based systems. The intrinsic dynamic range is typically of the order of 200–1000 (~8 to ~10 bits). Images of largely different intensity can be easily acquired by varying the gain of the MCP and/or the internal amplification stages of the camera. This can be done under computer control to obtain images of optimum brightness and contrast for each ion species [51]. However, in this way the direct relation between the ion intensity and the grey level values in the image is lost unless a sophisticated calibration scheme is set up. To overcome this problem, one can measure the total ion

current of the image with the electron multiplier or Faraday detector just after the acquisition of the image.

Within one image the dynamic range can be extended by real-time integrating the image and storing the 8-bit subimages which are not yet in overflow. This integration is possible because of the very low background of the camera (0.004 counts s^{-1} $pixel^{-1}$). In this way, a dynamic range of about 4000 is possible [51].

2.6.5.2 Influence of Mass Resolution

Ion microscope and ion microprobe images are often acquired at low mass resolution ($M/\Delta M \sim 300$). This resolution is sufficient to separate ions of different nominal mass; however, interference from molecular ions or organic fragments are very likely at this resolution. Figure 8a shows the secondary ion image at mass 194 acquired at low resolution with a Cameca 4f instrument taken from a Teflon-coated Pt–Ir wire with a diameter of 30 µm. The wire was cut at an angle so that the Pt–Ir core is visible. In Fig. 8b and c the same area is imaged at high mass resolution. Figure 8b corresponds to $^{194}Pt^+$ ions, and reveals the core of the wire, whereas Fig. 8c shows the distribution of an organic fragment due to the Teflon.

2.6.5.3 Image Sequences

Images of different ions are acquired sequentially with a magnetic sector instrument. The mass spectrometer is tuned to a certain mass, and the image is acquired (averaged or integrated). The mass spectrometer is then adjusted for the next predefined mass, and again an image is taken. The entire sequence takes up to several minutes, depending on the number of masses (typical 10–20) and the integration time per image. If image depth profiles are acquired, this cycle is repeated several times. An image depth profile can result in a very large amount of data; for example, if 50 cycles of 10 images each are taken, in total 500 images (of 256 × 256 pixels) need to be stored. Such an image depth profile is acquired in about 1 h of instrument time and, depending on the applied primary beam current and imaged field, corresponds to a sampled volume with an area of 100 × 100 μm^2 and a depth of 1 µm.

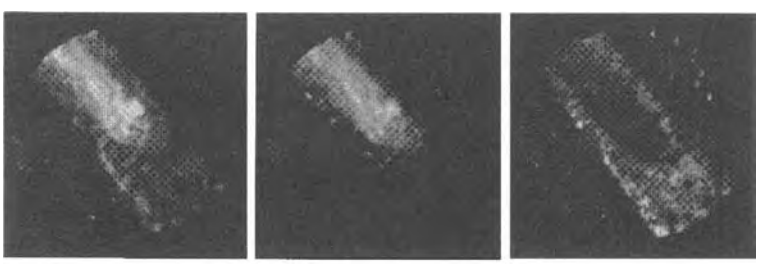


Figure 8. (a) Image at mass 194 of a cross-section of a Teflon-coated Pt–Ir wire acquired with the Cameca 4f ion microscope at low mass resolution. Images of the same area at high mass resolution showing the contribution from (b) the

2.6.5.4 Interpretation and Processing of Ion Images

Ideally the contrast observed in an ion image should be only due to local concentration variations of the measured ion. However, the relation between elemental concentration at the surface and the observed grey level in the acquired image is often disturbed by spurious contrast mechanisms. The main sources of spurious contrast are: matrix contrast, orientation or crystallographic contrast, topographic contrast, and chromatic or energy contrast [11]. The last two are less important in ion microprobe images. Finally, the uneven response of the position-sensitive ion detector used in the ion microscope (e.g., gain variations across the MCP) may cause unwanted intensity variations in the image. To interpret ion images correctly, the concentration contrast must be separated from the other contrast sources [6].

Crystallographic contrast occurs in polycrystalline materials because the sputter rate and thus the secondary ion intensity depend on the orientation of the crystallographic planes with respect to the incident primary ion beam. An example of this contrast is given in Fig. 9.

If the surface exhibits a certain roughness, topographic contrast can be observed. Due to the oblique incidence of the primary ion beam, some parts of the surface may not receive any primary ions. Also, the fact that the sputter yield depends on the incident angle contributes to this type of contrast.

Chromatic contrast is caused by the local variation in energy distribution of the secondary ions and the limited energy bandpass of the mass spectrometer. This contrast is most pronounced in the ion microscope, especially when charge build-up occurs on nonconducting samples.

Figure 9. Cu^+ secondary ion image of a pure copper sample, showing crystallographic contrast.

Matrix contrast causes the secondary ion yield of the element of interest to be different in different locations of the sample even if the concentration is the same, due to local variations in the composition of the matrix [52–54]. This false contrast is the most difficult to recognize, and may seriously affect the interpretation of the ion images.

Depending on the type of analysis, various image-processing steps might be required. Normalizing the image of a trace impurity to the image of a matrix element can be done to eliminate contrast due to uneven illumination or camera response and effects of crystal orientation. The normalized images then better reflect the contrast due to the concentration variation [55]. Truly quantitative images are more difficult to obtain because of the high variation in sensitivity and the influence

of the matrix on the sensitivity. The sensitivity factor and the so-called matrix/ion species ratio (MISR) have been applied successfully on a pixel basis to obtain images where each grey value corresponds to the local concentration of the element [51].

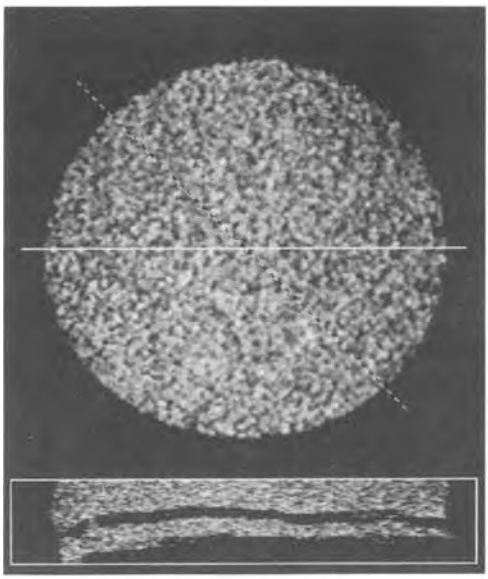
Classical image-processing procedures such as contrast enhancement or grey level equalizing and the use of pseudo-color look-up tables are often employed to aid the visual interpretation of the ion images [56]. Image restoration techniques using Fourier filtering have been used to enhance the lateral resolution of ion microprobe images [57, 58]. However, the high noise level in the ion microprobe images (pixels on average may contain only a few counts) limits the applicability of this technique.

2.6.5.5 Analysis of Image Depth Sequences

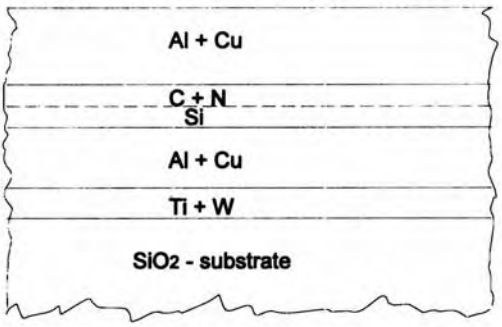
The acquisition of sequences of images as functions of depth allows a number of interesting data-processing options [59–66]. The images for one specific ion can be arranged in a stack, where the x and y directions correspond to the lateral distribution, and the z direction to depth. This stack can be probed in different ways. In local (or retro) depth profiling, an area is selected on top of the stack, and the intensity of the pixels in this area is calculated for the different images in the stack. The result is a depth profile (intensity versus depth). The interesting aspect is that this process can be repeated on different regions of the sample area without reanalyzing the sample. This is especially useful for the investigation of complex samples containing certain features buried below the surface. Cross-sections of the image stack allow the depth distribution of the various species to be viewed under different angles. Finally, a pseudo three-dimensional reconstruction of the interior structure can be made. Examples of these procedures are given below.

In the first example, the sample studied consisted of a 1.5 μm thick multilayer containing aluminum, copper, titanium, tungsten, and some other impurities on an SiO_2 substrate. An area of 150 μm in diameter was analyzed in the ion microscope mode with the Cameca 4f SIMS instrument. An image depth profile of 9 masses ($^{12}C^+$, $^{14}N^+$, $^{16}C^+$, $^{30}Si^+$, $^{49}Ti^+$, $^{27}Al_2^+$, $^{46}Ti^{16}O^+$, $^{63}Cu^+$, and $^{184}W^+$) and 44 cycles was acquired in about 90 min, resulting in 396 images of 256 × 256 pixels. Figure 10a shows the uniform distribution of aluminum at the surface. The full line indicates the location where a cross-section was made perpendicular to the surface. Figure 10b shows cross-sectional images through each of the nine image stacks. From these images one can clearly deduce the structure of the multilayer, which is schematically represented in Fig. 10c.

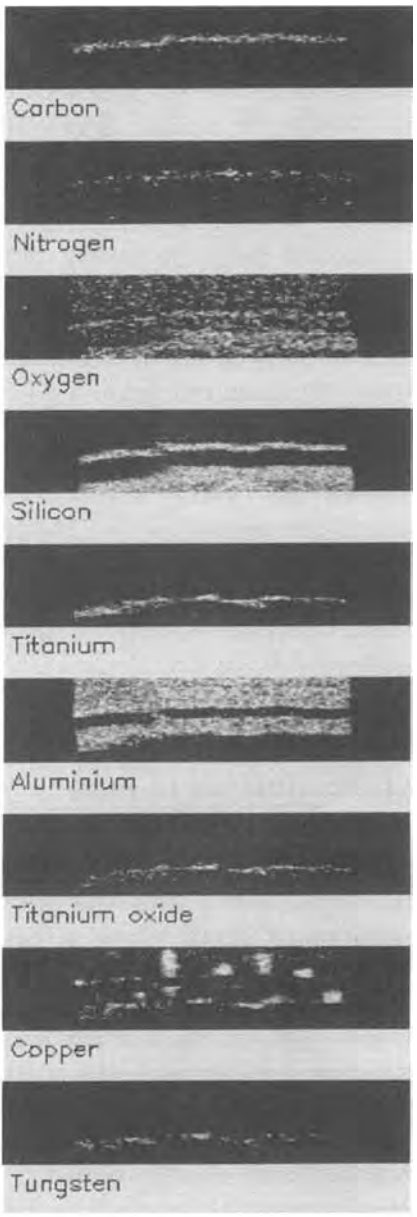
In the second example we consider the pseudo three-dimensional reconstruction of a sample consisting of aluminum structures (lanes and planes) deposited on an SiO_2 substrate and covered with an SiO_2 passivation layer. The passivation layer also contains aluminum inclusions. Thirty-five cycles of six masses ($^{11}B^+$, $^{23}Na^+$, $^{24}Mg^+$, $^{27}Al^+$, $^{31}P^+$, and $^{39}K^+$) were acquired under conditions similar to the first example. Figure 11 shows the lateral distribution of aluminum as a



a)



c)



b)

Figure 10. (a) Al^+ image at the surface of a multilayer structure. The full line indicates the position where a vertical cross-section through the image stack is made. (b) Cross-sectional images obtained from the image stacks of nine different ions. (c) Schematic view of the reconstructed multilayer.

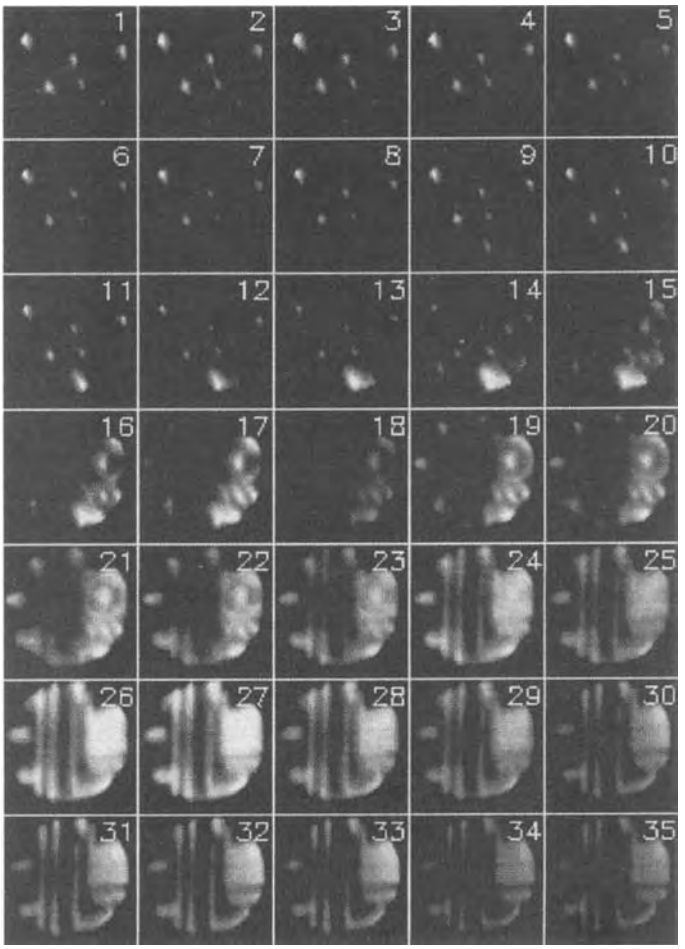


Figure 11. Lateral distribution of aluminum as a function of depth in a sample containing aluminum structures buried under an SiO_2 passivation layer. Cycle 1 corresponds to the surface, and cycle 35 to ~1 μm below the surface.

function of depth. Cycles 1 (at the surface) to 11 show the aluminum inclusions in the passivation layer. From cycle 16 onwards, the buried aluminum structure becomes visible. From the stack of aluminum images, a pseudo three-dimensional reconstruction was made, as shown in Fig. 12. In this transparent view the grey scale corresponds to depth in the sample. Some of the aluminum inclusions in the passivation layer are in contact with the buried aluminum structure, causing device failure.

2.6.5.6 Analysis of Multivariate Ion Images

Images of different masses taken at the surface or at a certain depth can be arranged as a multivariate image stack. In this multivariate image stack the x and y directions correspond to the lateral distribution, while the z direction is now a spectroscopic dimension, that is, each image holds the same lateral (spatial) distribution of a different ion species.

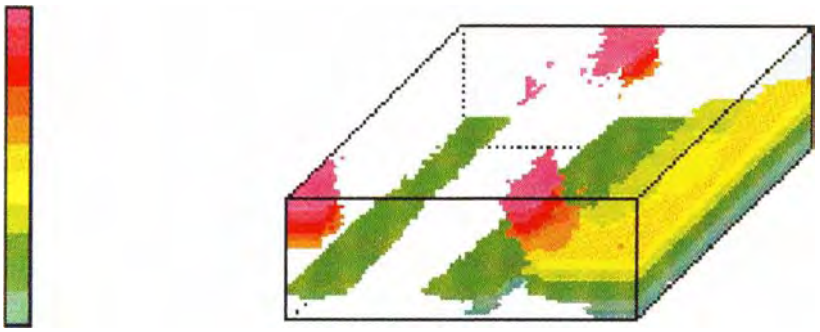


Figure 12. Reconstructed pseudo three-dimensional view of the aluminum distribution in the interior of a sample. In the transparent view the grey scale corresponds to depth in the sample.

Very often these images are correlated because species are present at the same location on the sample or are chemically related. Multivariate data analysis methods such as principal component analysis can be applied to this image stack [67–69]. The basic idea is to reduce the large number of correlated images to a smaller number of images (two or three) in such a way that the major portion of the information, or variation, present in all the images is condensed in these (principal component) images. This data reduction allows the user to observe the essential information. By making two-dimensional scatter plots of these first few principal component images, regions in the sample with similar chemical properties can be identified [70, 71]. This method is especially interesting for TOF-SIMS. The nature of the acquisition means that often a large number of images (20–100) are acquired simultaneously. Multivariate image analysis aids considerably in interpreting the large amount of data gathered.

As an example of the application of multivariate image analysis, we consider the study of a copper surface treated with butylbenzotriamide (BBT) [9]. BBT is a corrosion inhibitor forming a self-assembling monolayer on metals. The formation of the passivation layer is affected by the oxidation state of the metal surface and by the presence of impurities. Figure 13 shows four negative secondary ion images out of a series of 21 images in total. The images were acquired with a TOF SIMS instrument using a 1 μm diameter pulsed Ga^+ primary ion beam with a dose of 2.3×10^{12} ions cm^{-2}. In Fig. 13, mass 19 corresponds to F^-, mass 80 to SO_3^{2-}, mass 174 to deprotonated BBT $(BBT\text{-}H)^-$, and mass 411 to a Cu–BBT adduct $(Cu(BBT\text{-}H)_2)^-$. The other images at masses 12, 25, 32, 35, 38, 41, 42, 50, 81, 96, 97, 131, 143, 254, 255, 263, and 382 more or less resemble one of the four images shown.

Figure 14 shows the first three principal component images calculated from the 21 original images. The first two show a clear structure in the image, whereas the third (and following) principal component images contain mainly noise. When the two-dimensional histogram of the first two principal component images is made, three different regions, corresponding to pixels that have similar grey values, are observed (Fig. 15). Next, a mask is

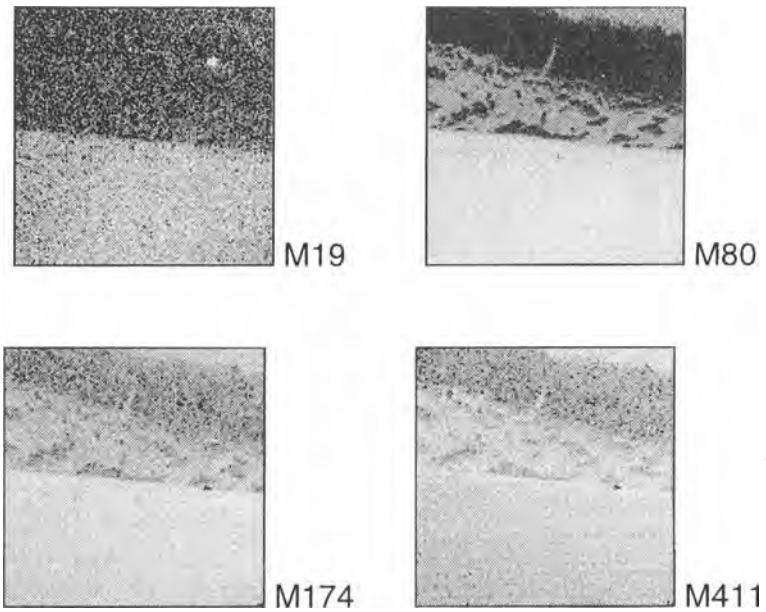


Figure 13. TOF-SIMS negative ion images from a copper surface treated with BBT: mass 19, F^-; mass 80, SO_3^{-2}; mass 174, $(BBT\text{-}H)^-$; mass 411, $(Cu(BBT\text{-}H)_2)^-$.

constructed with the pixels that contributed to each of these three regions. In Fig. 16 these three masks are shown. Mask A corresponds to class A in Fig. 15, and shows the area on the sample where the BBT molecule is associated with the copper surface. Mask B (class B in Fig. 15) corresponds to unreacted BBT, and mask C (class C in Fig. 15) shows the area where only inorganic acids and other contaminants are present at the surface.

In this way it is possible to identify locations on the surface where specific types of chemical reactions have occurred.

2.6.6 Sample Requirements

Because one bombards the surface with charged particles and also extracts charged

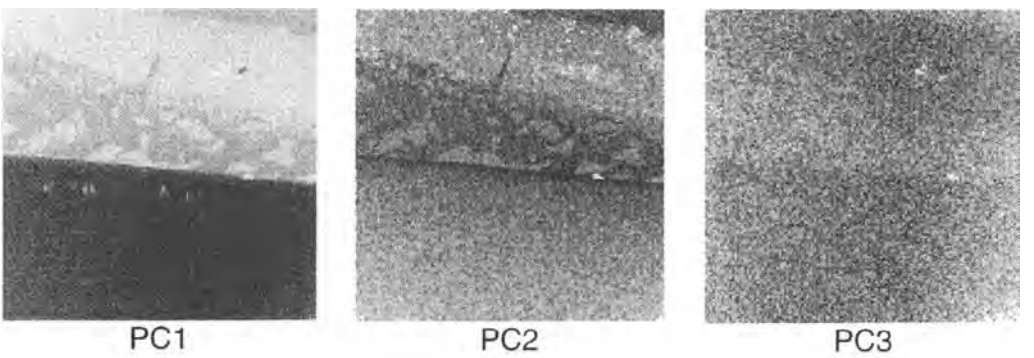


Figure 14. First three principal component images calculated from the 21 ion images from a BBT-treated copper surface.

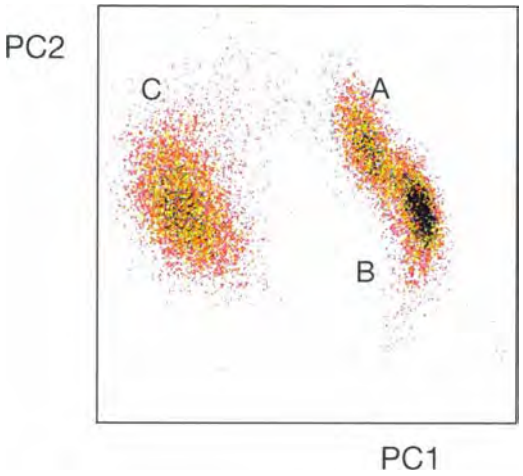


Figure 15. Two-dimensional histogram (scatter plot) of the first two principal component images shown in Fig. 14.

particles the samples should be conducting. Charge build-up will strongly affect the secondary ion extraction and cause considerable image degradation. Coating the sample with a conducting layer (gold), as is done in scanning electron microscopy (SEM), is less advantageous for SIMS. One has to sputter through the layer before reaching the sample surface. For true surface analysis this method is ruled out as the coating will destroy the surface composition. However, insulating samples can be analyzed when a positive primary ion beam is used and the charge build-up is compensated by flooding the surface at the same time with low energetic electrons.

Surface roughness is an important factor in SIMS and SIMS imaging. The sputter yield depends on the incident angle of the primary ions. Moreover, a shadowing effect occurs because the primary beam hits the sample typically at an angle of 40°, and some parts of the rough surface will not receive any primary ions. This causes a topographic induced contrast in the image. The roughness of the surface also degrades the depth resolution in depth profiling because it is assumed that all ions escape from the same depth, which is evidently not the case for a rough surface. The sample surface does not become smoother under continued bombardment; in fact, the roughness increases.

2.6.7 Application Domain

SIMS imaging is applied in many diverse fields of science and technology where the

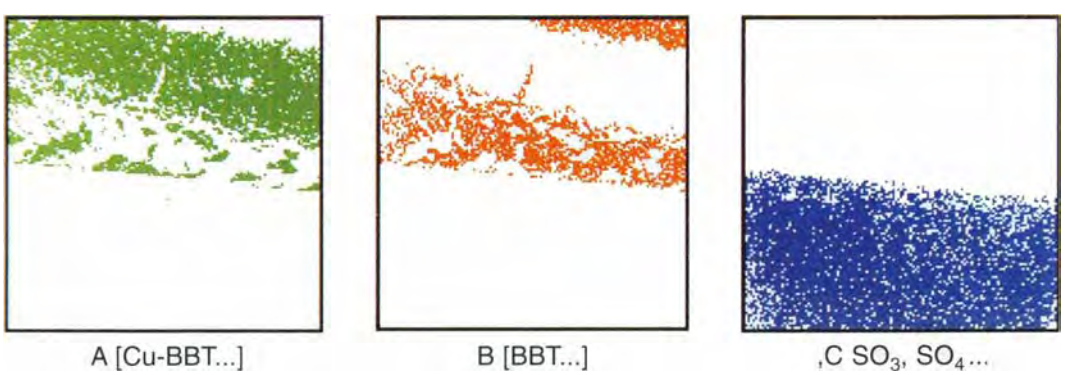


Figure 16. Masks obtained from the pixels that make up the three classes in Fig. 15. Masks A, B, and C correspond to classes A, B, and C, respectively, in Fig. 15.

more traditional SIMS methods are also commonly used: microelectronics, materials science [72, 73], geochemistry [74], biological and biomedical sciences [75, 76], and environmental studies. The imaging aspect of SIMS is of advantage whenever the spatial distribution of minor and trace elements needs to be investigated. In this respect, ion imaging complements SEM X-ray mapping with the added advantage of a much higher detection limit and the ability to perform three-dimensional analysis. Odom [49] provides a detailed overview of the use of imaging SIMS, with special emphasis on biological applications. The proceedings of the biannual SIMS conferences provide a good overview of the developments and applications of this imaging technique. It is interesting to note that at the 1982 SIMS conference about 8% of the contributions dealt with imaging aspects of SIMS whereas in 1991 this number had increased to more than 22%, clearly showing the growing importance of imaging.

Applications in microelectronics dealing with various aspects of manufacturing processes are reported in the literature: localization and identification of contaminants [77], study of device failure, three-dimensional characterization of implants, and the study of multilayer structures [78]. The ability of static SIMS to view the distribution of organic species at the surface is used to study contamination from cleaning solvents and other process chemicals employed in photolithography.

A considerable number of applications deal with the study of biological systems. Although this is probably the most difficult subject, the high sensitivity of imaging SIMS is of great value here, because biochemical processes are often controlled by a low concentration of chemical species localized in subcellular structures. For the analysis of soft tissues adequate sample preparation is necessary to maintain the cell structure and the chemical composition after removing or immobilizing the cell fluid. One possibility is to employ cryomicrotomy after shock freezing of the sample. If the SIMS instrument is equipped with a cryotransfer system and a cryo-stage, the section can be analyzed directly [79]. Alternatively, the section can be freeze dried under carefully controlled conditions. Other special sample preparation techniques have been developed. The work of Lodding [80, 81] gives a detailed account of the analysis of hard tissues such as teeth and bone.

SIMS imaging is also used to study particles and other microscopic objects. Atmospheric particulate material of a few micrometers in diameter can be studied on an individual particle basis to investigate the chemical composition and to disclose surface enrichments [82, 83]. Silver halide microcrystals used in photographic material are typically 1 μm in diameter and have a complex core–shell structure with different chemical compositions of the type AgX_xY_y where X and Y are Cl^-, Br^-, and/or I^-. SIMS imaging using the ion microprobe mode is used to reveal the internal structure of these microscopic crystals and to give feedback on the production process and on the photographic properties of the material [84].

A fast growing area of SIMS imaging is the mapping of organic coatings and layers using static TOF SIMS. The ability to localize organic species allows the verification of the uniformity of organic coatings on a wide variety of materials and

the study of the chemistry involved in adhesion [9].

2.6.8 References

[1] C. W. Magee, *Ultramicroscopy* **1984,** *14*, 55.
[2] W. Katz, G. A. Smith, *Scanning Electron Micros.* **1984,** 1557.
[3] K. T. F. Janssen, P. R. Boudewijn in *Analysis of Microelectronic Materials and Devices* (Eds.: M. Grasserbauer, H. W. Werner), Wiley, Chichester **1991,** p. 407.
[4] A. Balise, G. Slodzian, *Rev. Phys. Appl.* **1973,** *8*, 105.
[5] G. H. Morrison, G. Slodzian, *Anal. Chem.* **1975,** *47*, 933A.
[6] W. Steiger, F. G. Rüdenauer, *Anal. Chem.* **1979,** *51*, 2107.
[7] A. J. Patkin, B. K. Furman, G. H. Morrison in *Microbeam Analysis* (Ed.: D. B. Wittry), San Francisco Press, San Francisco **1980,** p. 181.
[8] F. G. Rüdenauer, *Surf. Interface Anal.* **1984,** *6*, 132.
[9] A. Benninghoven, *Angew. Chem., Int. Ed. Engl.* **1994,** *33*, 1023.
[10] G. Slodzian, *Scanning Microsc.* **1987,** 1.
[11] H. W. Werner in *Proc. Conf. on Electron and Ion Spectroscopy of Solids* (Eds.: L. Fiermans, J. Vennik, W. Dekeyser), Plenum Press, New York **1978,** p. 324.
[12] D. M. Drummer, G. H. Morrison, *Anal. Chem.* **1980,** *42*, 2147.
[13] S. R. Bryan, W. S. Woodward, R. W. Linton, D. P. Griffis, *J. Vac. Sci. Technol.* **1985,** *A3*, 2102.
[14] H. Hutter, M. Grasserbauer, *Mikrochim. Acta* **1992,** *107*, 137.
[15] A. Benninghoven, F. G. Rüdenauer, H. W. Werner (Eds.), *Secondary Ion Mass Spectrometry: Basic Concepts, Instrumental Aspects, Applications and Trends,* Wiley, New York **1987.**
[16] R. G. Wilson, F. A. Stevie, C. W. Magee, *Secondary Ion Mass Spectrometry,* Wiley, New York **1989.**
[17] P. Sigmund, *Phys. Rev.* **1969,** *184*, 383.
[18] V. R. Deline, C. A. Evans, *Appl. Phys. Lett.* **1978,** *33*, 578.
[19] C. A. Anderson, J. R. Hinthorn, *Anal. Chem.* **1973,** *45*, 1421.
[20] P. Williams, *Surf. Sci.* **1979,** *90*, 588.
[21] P. Williams, *Appl. Surf. Sci.* **1982,** *13*, 241.
[22] K. Wittmack, *Appl. Surf. Sci.* **1981,** *9*, 315.
[23] A. E. Morgan, H. A. M. de Grefte, H. J. Tolle, *J. Vac. Sci. Technol.* **1981,** *18*, 164.
[24] M. T. Bernius, Y.-C. Ling, G. H. Morrison, *Anal. Chem.* **1986,** *58*, 94.
[25] J. M. Chabala, R. Levi-Setti, Y. L. Wang, *Appl. Surface Sci.* **1988,** *32*, 10.
[26] R. Levi-Setti, G. Crow, Y. L. Wang in *Secondary Ion Mass Spectrometry V* (Eds.: A. Benninghoven, R. J. Colton, D. S. Simons, H. W. Werner), Springer-Verlag, Berlin **1986,** p. 132.
[27] H. A. Storms, K. F. Brown, J. D. Stein, *Anal. Chem.* **1977,** *49*, 2023.
[28] J. M. Chabala, R. Levi-Setti, Y. L. Wang, *J. Vac. Sci. Technol.* **1988,** *B6*, 910.
[29] M. Schuhmacher, H. N. Migeon, B. Rasser in *Secondary Ion Mass Spectrometry VIII* (Eds.: A. Benninghoven, K. T. F. Janssen, J. Tümpner, H. W. Werner), Wiley, Chichester **1992,** p. 49.
[30] J. C. Vickerman, A. Brown, N. M. Reed (Eds.), *Secondary Ion Mass Spectrometry: Principles and Applications*, Clarendon Press, Oxford **1989.**
[31] J. Schweiters, H. G. Cramer, U. Jürgens, E. Niehuis, T. Heller in *Secondary Ion Mass Spectrometry VIII* (Eds.: A. Benninghoven, K. T. F. Janssen, J. Tümpner, H. W. Werner), Wiley, Chichester **1992,** p. 497.
[32] J. D. Fasset, J. R. Roth, G. H. Morrison, *Anal. Chem.* **1977,** *49*, 2322.
[33] R. W. Odom, B. K. Furman, C. A. Evans, Jr., C. E. Bryson, W. A. Petersen, M. A. Kelly, D. H. Wayne, *Anal. Chem.* **1983,** *55*, 574.
[34] B. K. Furman, G. H. Morrison, *Anal. Chem.* **1980,** *52*, 2305.
[35] D. Newbury, D. Bright in *Microbeam Analysis* (Ed.: D. Newbury), San Francisco Press, San Francisco **1988,** p. 105.
[36] F. P. Michiels, W. K. Vanhoolst, P. E. Van Espen, F. C. Adams in *Microbeam Analysis* (Ed.: P. E. Russell), San Francisco Press, San Francisco **1989,** p. 594.
[37] J. D. Brown, *Scanning Microsc.* **1988,** *2*, 653.
[38] J. L. Hunter, S. F. Corcoran, R. W. Linton, D. P. Griffis in *Microbeam Analysis* (Ed.: P. E. Russell), San Francisco Press, San Francisco **1989,** p. 597.
[39] D. S. Mantus, G. H. Morrison, *Anal. Chem.* **1990,** *62*, 1148.
[40] H. N. Migeon, C. Le Pipec, J. J. Le Goux in *Secondary Ion Mass Spectrometry V* (Eds.: A. Benninghoven, R. J. Colton, D. S. Simons, H. W. Werner), Springer-Verlag, Berlin **1986,** p. 155.
[41] H. N. Migeon, B. Rasser, M. Schuhmacher, J. J. Le Goux in *Secondary Ion Mass Spectrometry VIII* (Eds.: A. Benninghoven, K. T. F. Janssen, J. Tümpner, H. W. Werner), Wiley, Chichester **1992,** p. 195.

[42] E. Niehuis in *Secondary Ion Mass Spectrometry VIII* (Eds.: A. Benninghoven, K. T. F. Janssen, J. Tümpner, H. W. Werner), Wiley, Chichester **1992,** p. 269.
[43] J. Schwieters, H. G. Cramer, T. Heller, U. Jürgens, E. Niehuis, J. Zehnpfenning, A. Benninghoven, *J. Vac. Sci. Technol.* **1991,** *A9*, 2864.
[44] R. Levi-Setti, G. Crow, Y. L. Wang, *Scanning Electron Microsc.* **1985,** 535.
[45] R. Levi-Setti, J. M. Chabala, K. Gavrilov, J. Li, K. K. Soni, R. Mogelevsky in *Secondary Ion Mass Spectrometry IX* (Eds.: A. Benninghoven, Y. Nikei, R. Shimuzi, H. W. Werner), Wiley, Chichester **1994,** p. 233.
[46] B. Schueler, P. Sander, D. A. Reed, *Vacuum* **1990,** *41*, 1661.
[47] B. Schueler in *Secondary Ion Mass Spectrometry VII* (Eds.: A. Benninghoven, C. A. Evans, Jr., K. D. McKeegan, H. A. Storms, H. W. Werner), Wiley, New York **1990,** p. 311.
[48] M. Schumacher, H. N. Migeon, B. Rasser in *Secondary Ion Mass Spectrometry VII* (Eds.: A. Benninghoven, C. A. Evans, Jr., K. D. McKeegan, H. A. Storms, H. W. Werner), Wiley, New York **1990,** p. 939.
[49] R. W. Odom in *Microscopic and Spectroscopic Imaging of the Chemical State* (Ed.: M. D. Morris), Marcel Dekker, New York **1993,** p. 345.
[50] R. W. Odom, D. H. Wayne, C. A. Evans in *Secondary Ion Mass Spectrometry IV* (Eds.: A. Benninghoven, J. Okano, R. Shimizu, H. W. Werner), Springer-Verlag, Berlin **1984,** p. 186.
[51] F. Michiels, W. Vanhoolst, P. Van Espen, F. Adams, *J. Am. Soc. Mass Spectrom.* **1990,** *1*, 37.
[52] J. T. Brenna, G. H. Morrison, *Anal. Chem.* **1986,** *58*, 1675.
[53] F. Michiels, F. Adams, *J. Anal. Atom. Spectrom.* **1987,** *2*, 773.
[54] R. Canteri, L. Moro, M. Anderle in *Secondary Ion Mass Spectrometry VIII* (Eds. A. Benninghoven, K. T. F. Janssen, J. Tümpner, H. W. Werner), Wiley, Chichester **1992,** p. 135.
[55] A. N. Thorne, A. Dubus, F. Degrève, *Scanning Electron Microsc.* **1986,** 1255.
[56] D. E. Newbury, D. S. Bright in *Secondary Ion Mass Spectrometry VII* (Eds.: A. Benninghoven, C. A. Evans, Jr., K. D. McKeegan, H. A. Storms, H. W. Werner), Wiley, New York **1990,** 929.
[57] B. G. M. Vandeginste, B. C. Kowalski, *Anal. Chem.* **1983,** *66*, 557.
[58] W. Hao, P. Van Espen, *Vacuum* **1994,** *45*, 447.
[59] F. Michiels, W. Vanhoolst, W. Jacob, P. Van Espen, F. Adams, *Micron Microsc. Acta* **1987,** *18*, 249.
[60] D. R. Kingham, A. R. Bayly, D. J. Fathers, P. Vohralic, J. M. Walls, A. R. Waugh, *Scanning Microsc.* **1987,** 463.
[61] J. J. Lee, K. H. Gray, W. J. Lin, J. L. Hunter, R. W. Linton in *Secondary Ion Mass Spectrometry VIII* (Eds.: A. Benninghoven, K. T. F. Janssen, J. Tümpner, H. W. Werner), Wiley, Chichester **1992,** p. 505.
[62] N. S. McIntyre, K. F. Taylor, G. R. Mount, C. G. Weisener in *Secondary Ion Mass Spectrometry VIII* (Eds.: A. Benninghoven, K. T. F. Janssen, J. Tümpner, H. W. Werner), Wiley, Chichester **1992,** p. 513.
[63] A. J. Patkin, G. H. Morrison, *Anal. Chem.* **1982,** *54*, 2.
[64] W. Steiger, F. G. Rüdenauer, G. Ernst, *Anal. Chem.* **1986,** *58*, 2037.
[65] G. Gillen, R. L. Myklebust in *Secondary Ion Mass Spectrometry VIII* (Eds.: A. Benninghoven, K. T. F. Janssen, J. Tümpner, H. W. Werner), Wiley, Chichester **1992,** 509.
[66] W. K. Vanhoolst, P. J. Van Espen, *Mikrochim. Acta* **1991,** 415.
[67] K. Esbensen, P. Geladi, *Chemometrics Intelligent Lab. Systems* **1989,** *7*, 67.
[68] P. Geladi, S. Wold, K. Esbensen, *Anal. Chim. Acta* **1986,** *191*, 473.
[69] P. Van Espen, G. Janssens, W. Vanhoolst, P. Geladi, *Analysis* **1992,** *20*, 81.
[70] N. Bonnet, *Mikrochim. Acta* **1995,** *120*, 195.
[71] N. Bonnet, M. Herbin, P. Vautrot, *Ultramicroscopy* **1995,** *60*, 349.
[72] D. B. Williams, R. Levi-Setti, J. M. Chabala, D. E. Newbury, *J. Microsc.* **1987,** *148*, 241.
[73] F. Adams, L. Butaye, G. Janssens, P. Van Espen, *Anal. Sci.* **1991,** *7*, 383.
[74] J. F. Lovering, *NBS Special Publication 427*, US National Bureau of Standards, Gaithersburg, MD **1975.**
[75] W. A. Ausserer, Y.-C. Ling, S. Chandra, G. H. Morrison, *Anal. Chem.* **1989,** *61*, 2690.
[76] C. M. John, J. A. Chakel, R. W. Odom in *Secondary Ion Mass Spectrometry VIII* (Eds.: A. Benninghoven, K. T. F. Janssen, J. Tümpner, H. W. Werner), Wiley, Chichester **1992,** p. 657.
[77] B. W. Schueler, R. W. Odom, J. A. Chakel in *Secondary Ion Mass Spectrometry VIII* (Eds.: A. Benninghoven, K. T. F. Janssen, J. Tümpner, H. W. Werner), Wiley, Chichester **1992,** p. 281.
[78] P. K. Chu, W. C. Harris, G. H. Morrison, *Anal. Chem.* **1982,** *54*, 2208.
[79] M. T. Bernius, S. Chandra, G. H. Morrison, *Rev. Sci. Instrum.* **1985,** *56*, 1347.
[80] A. Lodding, H. Odelius, *Mikrochim. Acta* **1983,** *10*, 21.
[81] A. Lodding, *Scanning Electron Microsc.* **1983,** 1229.

[82] G. E. Cabaniss, R. W. Linton, *Environ. Sci. Technol.* **1984,** *18*, 319.
[83] T. R. Keyser, D. F. S. Natusch, C. A. Evans, Jr., R. W. Linton, *Environ. Sci. Technol.* **1978,** *12*, 768.
[84] I. Geuens, R. Gijbels, W. Jacob in *Secondary Ion Mass Spectrometry VIII* (Eds.: A. Benninghoven, K. T. F. Janssen, J. Tümpner, H. W. Werner), Wiley, Chichester **1992,** p. 479.

Part V

Magnetic Methods

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

1 Nuclear Magnetic Resonance

1.1 Introduction

The aim of this Chapter is to present a physical picture of the basic principles of nuclear magnetic resonance (NMR) microscopy. Since the focus is on spatial information, we take the unusual approach of discussing the NMR experiment in terms of magnetization gratings, how they are created and how they evolve. For length scales that are longer than about 100 Å, spatial NMR measurements rely on the use of magnetic field gradients to create a spatial grating in the spin magnetization. The spatial resolution of the measurement depends on the pitch of this grating and the period for which it can be maintained. To acquire an image, however, also requires that each individual volume element, a voxel, contains a detectable number of nuclear spins. Since NMR is a relatively insensitive technique, voxels are typically much larger than the finest grating pitch that can be created. In certain cases spatial information is available at these higher resolutions by performing a form of scattering experiment. Figure 1 shows the spatial scales of NMR imaging and scattering.

In the imaging regime, each set of spins in the sample is correlated with a particular volume element, and the desired information is either the density of spins, or the density of a subset of the sample. In the scattering case the information is a local correlation, such as the mean distance that spins can move over a short period. For instance, if the displacement is molecular diffusion of a liquid confined to small regions by boundaries, then the scattering experiment will return the size distribution of the confining cells, but will include no information on the location of an individual cell. Since each set of spins in the sample contributes coherently to the overall signal, then the sensitivity of the measurement is independent of the length scale, and arbitrarily small features can be explored. In imaging where each feature is observed as an isolated response, this is no longer the case. In Fig. 1 the horizontal line at about 7 μm separates the area above which imaging is viable and below which scattering methods must be employed.

Image contrast is an important issue which we will not have the space to explore. An NMR spectrum is rich in chemical and morphological detail, and any feature (chemical shift, coupling, relaxation time) may be used as the basis for creating contrast in NMR imaging. Any imaging experiment may be preceded by a period of contrast creation where a select spectroscopic feature is enhanced relative to another. In such fashion a single

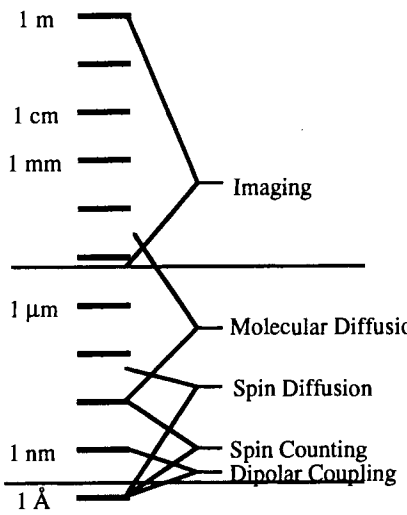


Figure 1. Spatial scales of NMR experiments. The scale is split into three regions, at the longest length scales absolute spatial information is recorded where the signal is correlated to a specific area of space. In the intermediate length scales a resolved volume element no longer contains sufficient number of spins to be individually resolved, but one can still measure average properties of the system, often associated with some form of transport, the scattering regime. In both of the above, linear magnetic field gradients are employed to create a well-defined spatial grating. At the highest resolutions, a grating can no longer be created via the gradient, and methods rely instead on spatially heterogeneous chemistry. Here, detailed average information such as internuclear distances is still available, but the information is not obtained in a linear fashion.

chemical component in a mixture may be selected as the basis for image formation, and in the solid state the selectivity may be extended to a specific molecular orientation. The details of such experiments have been reviewed, and a list of general references is included at the end of the chapter.

1.2 Background

In its most general form, the NMR experiment requires a quantum mechanical description which is usually in terms of the density matrix approach since the measured bulk magnetization is from a collection of nuclear spins. For non-interacting spin $= \frac{1}{2}$ nuclei, however, a semiclassical description introduced by Bloch provides a complete picture of the dynamics, and is adequate for most aspects of the imaging experiment. Only the Bloch picture will be introduced here.

Most elements have at least one isotope with a nonzero spin angular momentum, $\boldsymbol{I}$, and an associated magnetic moment, μ, which are related by the gyromagnetic ratio, γ,

$$\mu = \hbar\gamma \boldsymbol{I} \tag{1}$$

Both the gyromagnetic ratio and the magnitude of the nuclear spin I are characteristic of the nuclear state and are invariant. Protons have the highest gyromagnetic ratio of stable nuclei with $\gamma = 2\pi$ 4250 Hz G^{-1}, $I = \frac{1}{2}$, and are nearly 100% naturally abundant. For sensitivity reasons the majority of imaging experiments are measures of proton spin density.

The NMR experiment involves the interaction of the bulk magnetic moment, $\boldsymbol{M}$, of a collection of nuclear spins with a strong externally applied magnetic field. The interaction follows the Zeeman interaction, with the energy being proportional to the field, $\boldsymbol{B}_0$,

$$E = -\boldsymbol{M} \cdot \boldsymbol{B}_0 \tag{2}$$

The bulk magnetization is a superposition of the magnetic moments of a large number of individual spins. For spin $= \frac{1}{2}$ nuclei, the nuclear spin is a magnetic dipole with two allowed states, parallel and antiparallel to the applied field, with an energy difference of

$$\Delta E = \hbar\gamma|B_0| = \hbar\omega_0 \tag{3}$$

This energy difference is quite small, and the associated frequency of the electromagnetic photons, ω_0, is in the radio-frequency (RF) portion of the spectrum. A collection of spins in thermal equilibrium will populate these two states according to Boltzmann statistics,

$$\frac{N_-}{N_+} = e^{-\Delta E/kT} \tag{4}$$

At field strengths that can be created in superconducting magnets (1–10 T), and at room temperatures, the resulting population difference is about 1 spin in 10^6. The detected magnetization is this small excess, and so the NMR experiment is relatively insensitive; normally greater than 10^{15} spins are required for detection.

Since the individual spins are quantized along the direction of the external magnetic field (along the z-axis), then only this component of the nuclear spin has a definite value at equilibrium, the two transverse components (x and y) are in superposition states. The bulk nuclear magnetization at equilibrium is then a stationary magnetic moment aligned along the z-axis.

The dynamics of the bulk magnetization away from equilibrium can be broken down into two simple motions: (1) a precession about the applied magnetic field associated with the torque originating from Eq. (2).

$$\frac{d}{dt}\boldsymbol{M}(t) = \gamma \boldsymbol{M}(t) \times \boldsymbol{B} \tag{5}$$

and (2) a relaxation that carries the magnetization back to equilibrium.

Since the external field is about the z-axis, the precession is around the z-axis at a frequency of

$$\omega_0 = \gamma |\boldsymbol{B}_0| \tag{6}$$

the so-called Larmor frequency. This linear dependence of the precession frequency to the applied magnetic field will form the basis of NMR imaging methods.

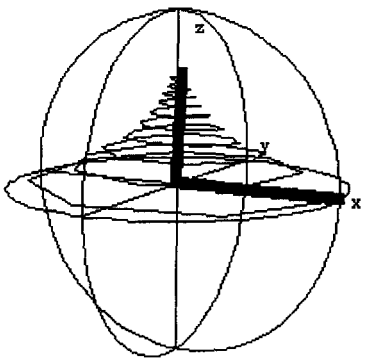
The relaxation process has two components, the magnetization along the z-axis relaxes towards its equilibrium value, M_0, and the transverse components relax towards zero. Since the bulk magnetization is composed of the magnetic moments of many individual spins, the length of the magnetization vector has a maximum but no minimum value (two individual magnetic moments can be antiparallel and thereby reduce the length of the bulk magnetization vector). Therefore, the transverse component may relax at a faster rate than the z-component does in its return to equilibrium, and two time constants are required to describe spin relaxation. These two relaxation times are called the spin–lattice relaxation time, T_1 (along the z-axis) and the spin–spin relaxation time, T_2 (governing the decay of transverse magnetization).

The Bloch equations describing the complete dynamics of a collection of isolated spins are

$$\frac{d}{dt}M(t) = \gamma M(t) \times B - \frac{1}{T_1}(M_z - M_0)\hat{z} - \frac{1}{T_2}(M_x\hat{x} + M_y\hat{y}) \tag{7}$$

Notice that T_2 must be less than or equal to T_1 to conserve the maximum length of the magnetization vector. Two useful pictorial representations of the evolution of the spin magnetization back to equilibrium, after having first been placed along the x-axis, are shown in Fig. 2.

In terms of the Bloch equations, Eq. (7), notice that two types of magnetic fields can



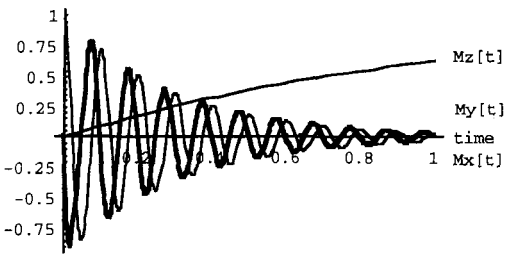


Figure 2. Diagrams of the return of the spin magnetization to equilibrium after being aligned along the x-axis. In both figures the evolution of a single bulk magnetization vector is being followed. The initial position is shown as the transverse vector at the top, which spirals in to the z-axis, the vertical vector. In the lower figure the three individual components of the magnetization are shown as a function of time. The NMR experiment measures the two transverse components, M_x and M_y. There are three motions, a precession about the z-axis, a decay of the transverse components and a slower growth along z toward the static equilibrium value.

have a pronounced influence on the dynamics: static fields along the z-axis, and transverse fields rotating at the Larmor frequency. Changes in the strength of the field along the z-axis result in changes in the Larmor frequency. Since the magnetization is precessing at the Larmor frequency, transverse magnetic fields that rotate about z at the same frequency (resonant fields) will create a constant torque on the spins and rotate them away from the z-axis. This second motion is most readily seen in a reference frame where the magnetization is stationary, the 'rotating' frame defined by the transformation

$$\frac{\mathrm{d}}{\mathrm{d}t}M_{\text{rotating}} = \frac{\mathrm{d}}{\mathrm{d}t}M_{\text{lab}} - \omega_0 M_x \hat{y} + \omega_0 M_y \hat{x} \tag{8}$$

If the applied magnetic fields are separated into two components, a static field along the z-axis, B_0, and a rotating field along the x-axis of the rotating frame, B_1, then the Bloch equations simplify to

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}M_x &= \Delta\omega M_y - \frac{M_x}{T_2} \\
\frac{\mathrm{d}}{\mathrm{d}t}M_y &= -\Delta\omega M_x + \omega_1(t) M_z - \frac{M_y}{T_2} \\
\frac{\mathrm{d}}{\mathrm{d}t}M_z &= -\omega_1(t) M_y - \frac{M_z - M_0}{T_1}
\end{aligned} \tag{9}$$

where $\Delta\omega$ is a small off-resonance term (the difference between the rotating frame frequency and the Larmor frequency), and ω_1 is the strength of the rotating field (the RF field),

$$\omega_1 = \gamma |\boldsymbol{B}_1| \tag{10}$$

For simplicity, the NMR experiment is usually referenced to a frame that moves with an externally applied RF field.

Notice that in transforming to the rotating frame, only those components that will influence the spin dynamics were kept, the fields that are static in the rotating frame. All other fields will be purely oscillatory in this reference frame, and the long time behavior of the spins will not be influenced by them.

The overall NMR experiment can be described with reference to Eq. (9). The spins are placed in a homogeneous

magnetic field, and after the populations have reached equilibrium, an RF field is applied for a short time. This RF pulse rotates the bulk magnetization away from the z-axis at an angular velocity of ω_1. By varying the length of the RF pulse, the angle can be made equal to 90°, $\omega_1 t = \pi/2$ (a so-called $\pi/2$ pulse), and the magnetization will be aligned along the y-axis. The RF field is then turned off and the spins return to equilibrium. The transverse components of the magnetization vector (M_x and M_y in Fig. 2) are captured by monitoring the voltage induced in a turned coil wrapped around the sample.

The measured signal, the free induction decay (FID), is in the time domain, and according to Eq. (9) has the form

$$M_{x,y}(t) = M_0 \, e^{-i\Delta\omega t} \, e^{-t/T_2} \tag{11}$$

The Fourier transformation of the FID is a Lorentzian with a width of

$$\nu_{\text{fwhm}} = \frac{1}{\pi T_2} \tag{12}$$

where 'fwhm' denotes full width at half-maximum. Since both the Fourier transformation and the Bloch equations in the absence of an RF field are linear, we are free to break the NMR experiment into a superposition of identical experiments on each NMR distinct set of spins. So a general result is a frequency spectrum showing many resonances each corresponding to an identifiable spin system. The frequency differences may arise from small chemical shifts associated with the partial screening of the magnetic field by the electrons, from spin–spin couplings, or from local variations in the magnetic field strength due to the heterogeneity of the magnetic susceptibility throughout the sample.

1.3 Magnetic Field Gradients, Magnetization Gratings, and *k*-Space

In order to record an image of a sample, there must be a measurable difference introduced between two locations in the sample. The most straightforward approach to this is to apply a magnetic field gradient,

$$B = B_0 + \frac{\partial}{\partial u} B_z \tag{13}$$

where u is x, y, or z, so that the resonance frequency varies across the sample. Only the three partial derivatives of the z-component of the magnetic field are of interest, since the others correspond to static fields in the transverse direction of the laboratory frame, and thus rotating fields in the rotating frame that do not influence the dynamics.

The presence of a magnetic field gradient introduces a spatial heterogeneity into the experiment, and the observed bulk magnetization is that integrated over the entire sample. If a z-gradient ($\partial B_z/\partial z$) is applied to a sample whose spin density is described by $\rho(x, y, z)$ then the FID, Eq. (11), becomes

$$M_{x,y}(t) = M_0 \, e^{i\Delta\omega t} \, e^{-t/T_2} \times \int \rho(x, y, z) \times e^{-i\gamma(\partial B_z/\partial z)zt} \, dx \, dy \, dz \tag{14}$$

This is put in a more recognizable form by introducing the reciprocal space vector,

$$k_i = \int \gamma \frac{\partial}{\partial i} B_z(t) \, dt \tag{15}$$

and by defining the projection of the spin density along the z-axis,

$$P(z) = \int \rho(x, y, z)\, \mathrm{d}x\, \mathrm{d}y \tag{16}$$

so that

$$M_{x,y}(t, k_z) = M_{x,y}(t, 0) \int P(z)\, \mathrm{e}^{-\mathrm{i}k_z z}\, \mathrm{d}z \tag{17}$$

where $M_{x,y}(t, 0)$ is the NMR signal in the absence of the magnetic field gradient. The measured NMR signal in the presence of a magnetic field gradient is the Fourier transform of the projection of the spin density along the direction of the gradient convoluted with the NMR lineshape in the absence of the gradient,

$$\text{Image} = \text{ObjectFunction} \otimes \text{NMR lineshape} \tag{18}$$

It is helpful to understand how it is that the NMR signal is a measure of the Fourier components of the spin density. As shown in Fig. 3, spin evolution in the presence of a magnetic field gradient introduces a sinusoidal magnetization grating across the sample, and the $\boldsymbol{k}$-vector describes the spacing of the grating,

$$k_z = \frac{2\pi}{\lambda_z} \tag{19}$$

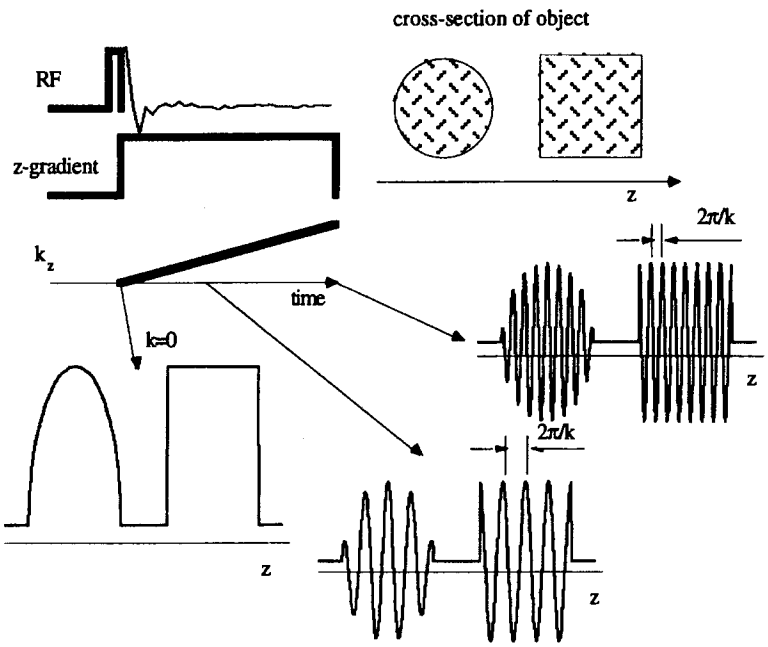


Figure 3. A schematic description of a one-dimensional NMR imaging experiment to measure the projection of the object spin density along a given axis. The object is shown and consists of two regions of spins with circular and square cross-sections. The z-axis is identified and the applied magnetic field gradient is a $\partial B_z/\partial z$ gradient, so the Larmor frequency of the spins in the object increases from left to right. The measurement is proceeded by an RF pulse that excites the spins and places them in the transverse plane. The three profiles of the object show one axis of the transverse magnetization at three different values of k. Higher k-values correspond to finer magnetization gratings. The NMR signal is the integrated spin magnetization across the sample, which for this sample has the shape shown following the RF pulse.

where λ_z is the period of the grating. The essence of the NMR imaging experiment is that by a suitable combination of magnetic field gradients oriented along the x, y, and z laboratory frame directions any Fourier component of the spin density can be measured. Since these gradients are under experimental control and can be switched on in a very short time compared to the NMR relaxation times, then there is a great deal of flexibility in how one proceeds to record all of the necessary Fourier components needed to reconstruct an image. We will return to this after discussing the limits to resolution, and multiple-pulse experiments.

In Eqs. (17) and (18), the NMR image is described as a convolution of a spatial function describing the density of the spins, and a frequency function describing the NMR spectrum. To make sense of this, the NMR spectrum must be converted into spatial units. Notice in Fig. 3 that in the presence of a constant magnetic field gradient, the two reciprocal space vectors, time and k, are directly proportional to each other,

$$\frac{k_z}{t} = \gamma \frac{\partial}{\partial z} B_z = \frac{\Delta\omega}{\Delta z} \tag{20}$$

So if the NMR spectrum corresponds to a single resonance line (such is the case for water) then the image resolution can be described in terms of the linewidth, ν_{fwhm} (or the spin–spin relaxation time), and the gradient strength,

$$\text{resolution} \geqslant \nu_{\text{fwhm}} \left(\gamma \frac{\partial}{\partial z} B_z \right)^{-1} \tag{21}$$

Sharper lines and stronger gradients lead to higher resolution. This is what makes NMR imaging of solids challenging since the T_2 is approximately 1000 times shorter for solids than for liquids.

There are other limitations to the obtainable resolution, the most important in microscopy of liquids being the random molecular motions of the spins. Water has a diffusion coefficient of about $3\,\mu m^2\,ms^{-1}$ at room temperature, and as the molecules move they carry their magnetic moment with them. This leads to an irreversible blurring of the magnetization grating and hence a loss in high spatial frequency information (Fig. 4). The influence of the spin motion on the magnetization grating is included in the Bloch equation by adding a diffusion term,

$$\frac{dM_i(t)}{dt} = D\nabla^2 M_i(t) \tag{22}$$

Equation (22) has solutions of

$$M_i(t) = M_i(0)\, e^{-Dk^2 t} \tag{23}$$

in a constant grating, and

$$M_{x,y}(k, t) = M_{x,y}(0)\, e^{-ikz}\, e^{-Dk^2 t/3} \tag{24}$$

in a constant magnetic field gradient. The factor of three in the exponential of Eq. (24) arises since the spacing of the grating is varying with time, and the attenuation is most pronounced for finer gratings.

Recognizing that there are potentially many contributions to the overall resolution of the imaging experiment, the convolution of Eq. (18) may be generalized as an ideal linear model,

$$\text{Image} = \text{ObjectFunction} \otimes \text{PointSpreadFunction} + \text{noise} \tag{25}$$

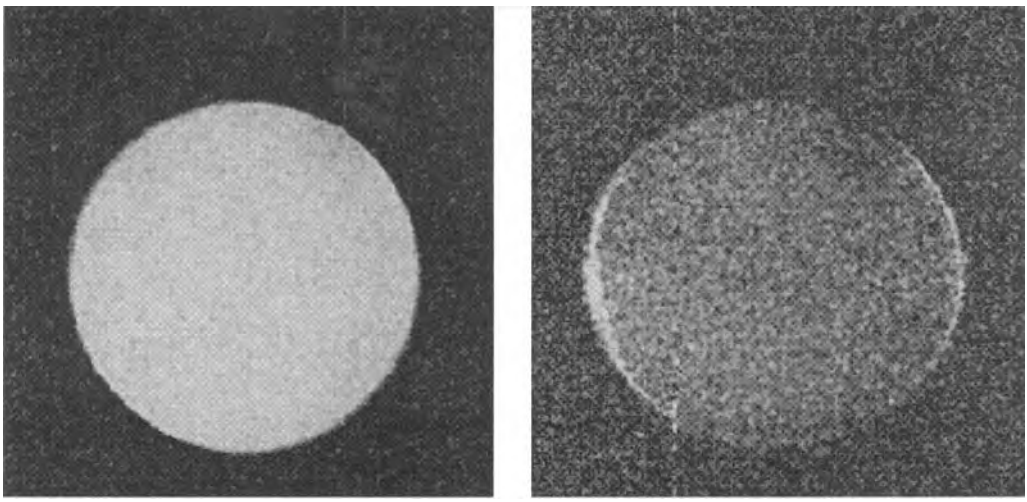


Figure 4. Two NMR images of a small tube of water. The image on the left was acquired with strong gradients so that molecular diffusion does not influence the image intensity, that on the right was acquired with much lower gradients strength and molecular diffusion destroyed the grating prior to acquisition. Notice the sharp half moons on the edges of the image; these arise from molecules that have an effective lower diffusion coefficient since they are near boundaries. The images are most sensitive to diffusion along the horizontal axis where a gradient echo was employed. The image field of view is 1.7 × 1.7 mm and the resolution is 6.6 × 6.6 μm.

The total point spread function (PSF) is then the convolution of the individual contributions, some of which are listed in Table 1.

In Table 1, interrupted molecular diffusion, which occurs where boundaries effectively reduce the diffusion coefficient, leads to a spatially dependent diffusion constant. Any contribution to the NMR linewidth that is spatially varying is nonlinear.

One of the reasons that NMR imaging is so successful is a result of the truncation of Hamiltonians to their secular components, that is, all of the interactions (except RF pulses) commute and hence do not interfere with one another. Since each interaction is itself linear, then the entire spin dynamics are linear and, to a good approximation, the linear model is accurate to the extent that the instrument is correctly engineered.

Table 1. Contributions to the NMR resolution

Contribution	Origin	Linear/nonlinear
NMR linewidth	T_2 relaxation	Linear
Bulk susceptibility	Variation of the local magnetic field due to variations in the local bulk susceptibility	Nonlinear
Free molecular diffusion	Random motion of spins	Linear
Interrupted molecular diffusion	Reduction of the effective diffusion coefficient by boundaries	Nonlinear
Instrumental imperfections	Nonlinearity in the gradients, inhomogeneity of the RF field	Nonlinear

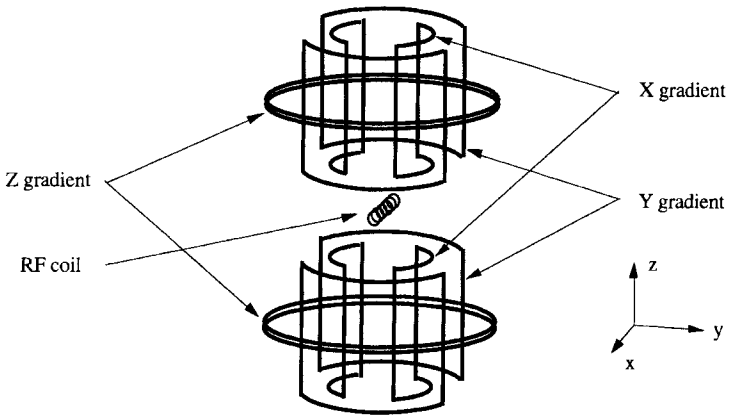


Figure 5. A schematic representation of the NMR microscopy probe; the magnetic field is along the vertical axis. The probe has a small RF coil wrapped tightly about the sample; the good filling of the coil is necessary to reduce resistive losses in the coil. This is surrounded by three gradient coils that are connected to audio-frequency amplifiers. The z-gradient is a Maxwell pair through which currents flow in opposite directions. Moving inwards, the x- and y-gradient coils are each composed of four semicircular current paths through which currents flow in parallel.

1.4 Nuclear Magnetic Resonance

There are many excellent descriptions of NMR instrumentation; here the focus is on the changes necessary to implement microscopy on a high-resolution, high-field spectrometer. Fortunately these are quite modest, since many spectrometers have the necessary gradient amplifiers and controllers, and all modern instruments are capable of performing the experiments. The fundamental difference is the NMR probe, shown schematically in Fig. 5.

Notice that the field of view is normally limited physically by the size of sample that can be accommodated in the RF coil. This is advantageous for sensitivity reasons, and quite necessary for reasonable experimental times. Some typical values are given in Table 2.

1.5 Echoes and Multiple-Pulse Experiments

That the various interactions, the gradient, chemical shift, susceptibility, and couplings, each drive the spin dynamics individually is useful and permits a simple linear model of the sensitivity and resolution of the imaging system, but equally important

Table 2 Some typical NMR microscopy probe configurations

Field of view	2.5 cm	1.0 cm	2.5 mm
Resolution	100 μm	20 μm	5 μm
Gradient strength	50 G cm^{-1}	100 G cm^{-1}	1000 G cm^{-1}
Magnet	9.6 T/89 mm	9.6 T/89 mm	14 T/54 mm

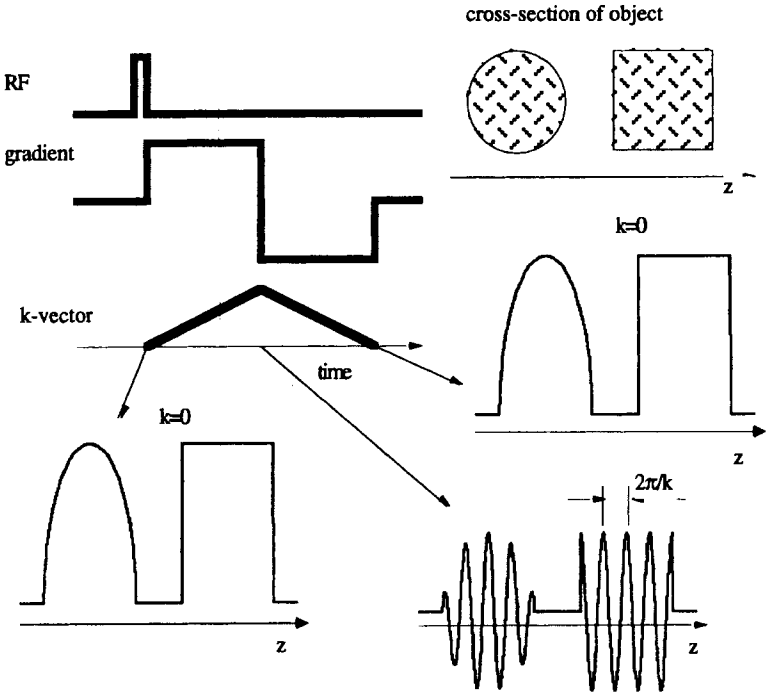


Figure 6. A gradient echo is generated by a bipolar gradient waveform. Since the k-vector is the integrated area under the gradient waveform, Eq. (15), at the end of the bipolar gradient k returns to zero, and there is no magnetization grating across the sample. At the mid-point of the gradient waveform k is at a maximum and the grating is at its finest. Since the NMR signal is the integration of the spin magnetization across the sample, the signal maxima correspond to the two points where k is zero, and hence an echo is observed.

(particularly in creating contrast) is the experimenter's ability to separate these by selectively refocusing one or more interactions. The gradient echo is perhaps the simplest example of this. Recall that the periodicity of the magnetization grating is the zero moment of the time-dependent gradient waveform (the integrated area of the waveform), Eq. (15). By applying the gradient as a bipolar pulse pair (Fig. 6) the grating is removed from the object at the end of the gradient waveform and the spins are refocused (back in phase). Since the only interaction that is influenced by the presence of a gradient is the gradient evolution itself, the bipolar gradient waveform shown in Fig. 6 will not refocus the chemical shift, or any other internal interaction. We will leave the applications of gradient echoes to later and turn to the question of how other interactions can be refocused.

In select cases the sign of the interaction can be changed, as it was in the gradient echo, but, in general, refocusing is accomplished through the actions of RF pulses that strongly modulate the spin dynamics. RF pulses are normally arranged such that the RF field strength is stronger than any internal interaction and so they can be thought of as instantaneously changing the spin state of the system.

The simplest RF pulse-created echo is the two-pulse $(\pi/2, \pi)$ sequence shown in Fig. 7. The first RF pulse rotates the magnetization from the z-axis into the transverse plane, where in the presence of a gradient it creates a magnetization grating. The π-pulse inverts the sense (directionality) of this grating so that continued evolution in the magnetic field gradient leads to the formation of an echo.

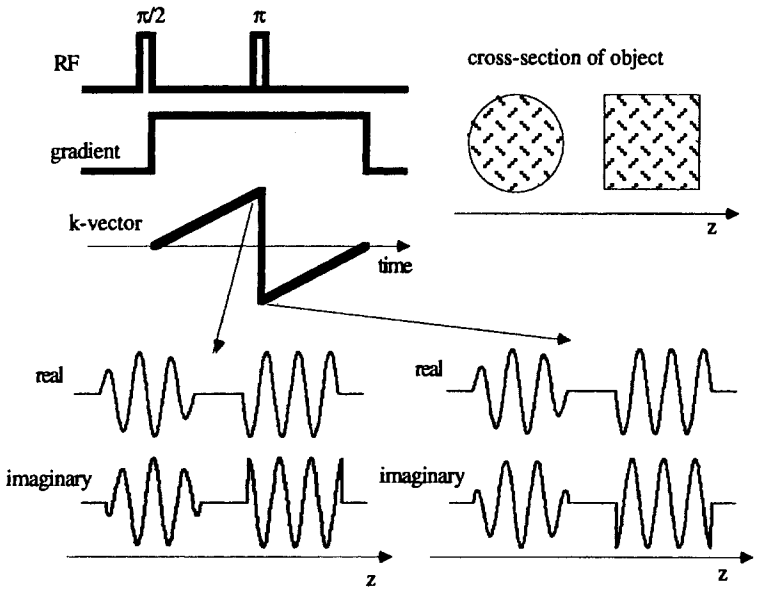


Figure 7. The figure shows the action of a π-pulse when applied to a magnetization grating. Here both quadrature components (real and imaginary) of the transverse magnetization must be kept track of. The two gratings are shown just before and just after the π-pulse which inverts the sign of one of the two components (the real component in the figure). This is consistent with the RF pulse corresponding to a rotation of 180° about an axis in the transverse plane. Before and after the π-pulse the magnetization grating has the same pitch, but the sign of the grating is changed by the pulse. Following the π-pulse the magnetic field gradient is still in the same direction and so the k-vector continues to increase, but now towards zero, and following an equal length of time k returns to zero and all of the magnetization is again coherent. A spin echo is formed.

The spin echo depicted in Fig. 7 will refocus any time-independent variations in the resonance frequency, including chemical shifts, susceptibility shifts, and heteronuclear spin couplings. Earlier the spin–spin relaxation time was identified with the decay of the bulk magnetization in the transverse plan, and we now see that this is not a fundamental relaxation phenomenon since, for example, the application of a magnetic field gradient can increase its rate, and since the magnetization can be refocused. The fundamental spin–spin relaxation rate is associated with those decay processes that cannot be refocused by methods such as the spin echo, and the observed decay of the transverse magnetization is the rate sum of this and other contributions,

$$\frac{1}{T_2^{\text{observed}}} = \frac{1}{T_2} + \frac{1}{T_2^{\text{other}}} \tag{26}$$

One very useful imaging application of the spin echo is to change the image contrast from being a measure of the pure spin density (by using a spin echo to refocus spatial variations in the observed T_2) to being dependent on the local magnetic field variations (by using a gradient echo).

Any train of RF pulse will create a pattern of echoes that can refocus line-broadening terms. A second application of spin echoes is to store information that can then be read out later. The stimulated echo shown in Fig. 8 is a classic example of this.

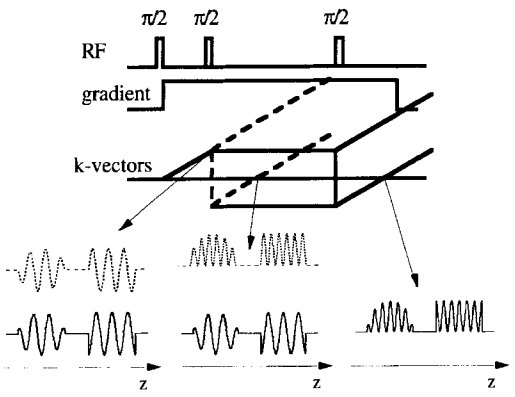


Figure 8. Schematic view of a $\pi/2$, $\pi/2$ echo and the stimulated echo. The second $\pi/2$ RF pulse rotates one component of the transverse spin magnetization to the z-axis (along the field), where it does not evolve (recall from the Bloch equation, Eq. (9), that the z-magnetization only slowly decays). The remaining transverse magnetization is now an oscillator function of the spatial coordinate and so is a combination of $+k$ and $-k$ states (the dotted trajectories in the figure). Likewise, the z-magnetization is also a combination of $+k$ and $-k$ (the solid lines). Between the second and third RF pulses, only the transverse components evolve, and only one-half of the signal forms an echo. The third RF pulse returns the magnetization that was stored along the z-axis back into the transverse plane, where it too evolves and eventually yields the 'stimulated' echo. This echo has special importance, since between the second and third pulses the corresponding magnetization was stored as a special grating.

One of the most useful features of the stimulated echo is that a magnetization grating can be stored for a time comparable to the spin–lattice relaxation time (on the order of a few seconds), which is an ideal approach to measuring extremely slow processes—for instance molecular transportation over macroscopic distances.

1.6 Two-Dimensional Imaging

Having seen that by applying a magnetic field gradient one can encode a one-dimensional image and collect a projection of the spin density of the sample, one clear path to two-dimensional imaging is to collect a series of such projections with the gradient at various orientations and then to use Radon-filtered back projection to reconstruct the image. Early NMR images were acquired in this fashion, and occasionally solid state images still are. However, most imaging is performed via Fourier imaging where a two- or three-dimensional region of reciprocal space is sampled corresponding to the desired resolution and field of view. Such data are easily measured since the gradients are under experimental control and virtually any gradient waveform can be generated. Since the various sequences can be rather complex, often a pictorial representation of reciprocal space is employed where the trajectory of the experiment is mapped out. It is then possible to focus on the manner in which two-dimensional k-space is sampled rather than to get caught up in the details of the NMR experiment.

Notice that the imaging experiment contains two fundamentally different times, a phase-encoding interval (t_1 in Fig. 9 during which no data are acquired) and a frequency-encoding interval (during data acquisition). These two times permit the separate encoding of the two interactions—the gradients in the two orthogonal directions—and thus allow all of k-space to be sampled. The image is the two-dimensional Fourier transform of the collected data.

The gradient evolution during the phase-encoding interval can be carried out in two fashions, the gradient strength can be kept constant and the time incremented, or the time can be kept constant

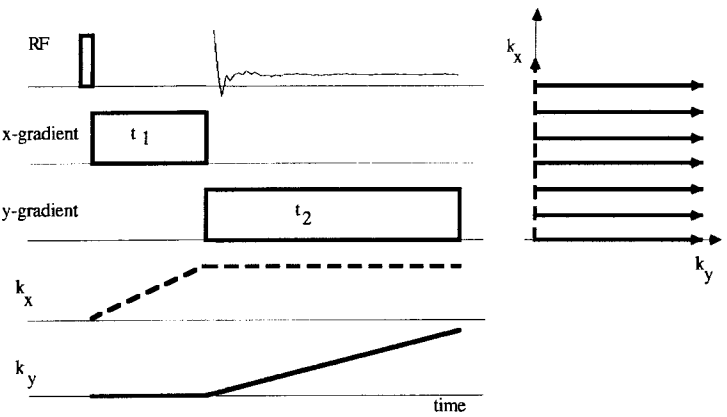


Figure 9. A generic two-dimensional Fourier imaging experiment. The data are collected on a Cartesian raster during the time period labeled t_2, and in the presence of a y-gradient. Prior to this, a brief x-gradient pulse has been applied to create a magnetization grating in the x-direction. The collection period therefore corresponds to the rays in the reciprocal space picture on the right, and each individual ray is collected during a separate experiment. The length of t_1 is systematically varied to achieve the desired offset along the k_x axis.

and the gradient strength incremented. The second, called spin warp imaging, has advantages since the extent of the evolution due to chemical shift or other non-gradient interactions is kept constant and these then appear solely as a signal attenuation factor. The resultant point spread function for a constant encoding time is an attenuated delta function (neglecting the contribution from sampling).

There is a wide range of imaging sequences, and we will not attempt to review these here; they all include the general features shown in Fig. 9. The two-dimensional experiments can be extended to three dimensions by encoding the third direction as a second phase-encoded axis.

1.7 Slice Selection

Recording a full three-dimensional image is often not the most economical approach to imaging, and slice selection can be achieved by taking advantage of the frequency offset dependence of RF excitation. Looking back at the Bloch equations, Eq. (9), for an on-resonance RF pulse ($\Delta\omega = 0$) the RF field is along the x-axis and hence the evolution of the spins is a simple rotation about the x-axis. However, as $\Delta\omega$ increases, then the dynamics become more complex and are most easily visualized by considering an 'effective' field that is the vector sum of the RF field along the x-axis and the off-resonance field along the z-axis. The dynamics are still a simple rotation about this effective field, but the motion of the magnetization vector now describes a cone rather than a plane. The result is that as the frequency offset is increased, the angle of the effective field to the z-axis decreases, and eventually the RF pulse has very little influence. The key to slice selection then is to apply a relatively weak RF pulse in the presence of a strong gradient, so that the frequency offset is spatially dependent (Fig. 10).

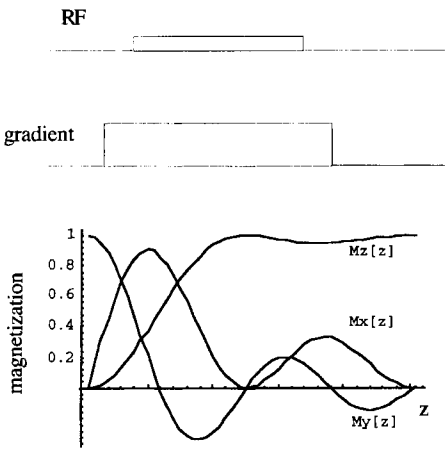


Figure 10. Calculated selective excitation profiles for a weak RF pulse in the presence of a magnetic field gradient. The pulse length is set to rotate the on-resonance spins (at the origin) by 90°. Notice that as the resonance offset increases (with increasing z), the effective rotation angle becomes smaller and most of the magnetization remains along the z-axis. For most images, a shaped RF pulse is employed that creates a square magnetization profile.

There is not a simple linear picture of the dynamics, but various shaped (amplitude-modulated) RF pulses have been developed that give well-defined square slice selection profiles.

So the overall slice-selected two-dimensional imaging experiment might look like that shown in Fig. 11.

Figure 12 shows images of rat arteries acquired using a multislice two-dimensional spin echo technique. Such images might form the basis of rapid, nondestructive, three-dimensional imaging.

1.8 Gratings and Molecular Motions

We have already seen in the case of molecular diffusion that molecular motions vary the magnetization grating. Random motions, such as diffusion, lead to an attenuation of the grating, and coherent processes such as flow displace the grating. The precession due to a z-gradient at a position z_0 has an instantaneous phase angle of $k_z z_0$, if the spins are not moving. In the presence of motion this angle (for a single spin packet) has the general form

$$\Theta_{\text{gradient}} = \gamma \int \frac{\mathrm{d}}{\mathrm{d}z} B_z(t) z(t)\, \mathrm{d}t \qquad (27)$$

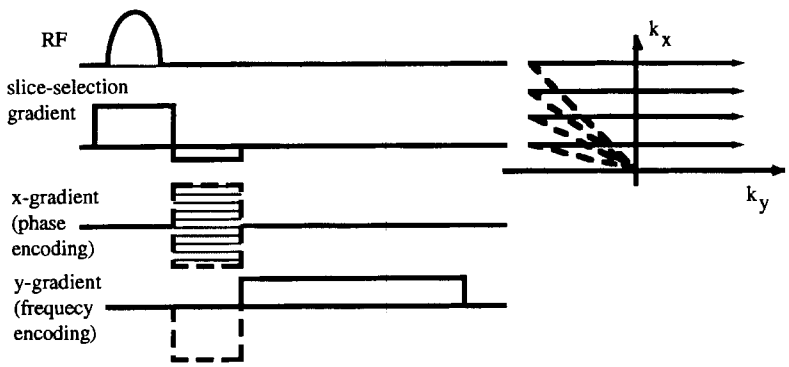


Figure 11. A slice-selected two-dimensional spin warp, gradient echo sequence. Notice that during the phase-encoding time the k_y vector is offset so that both positive and negative values can be sampled. Since NMR is a coherent spectroscopic technique, this has the advantage of measuring the phase. The gradient echo in the slice selection direction refocuses the evolution of the spins during the selective RF pulse.

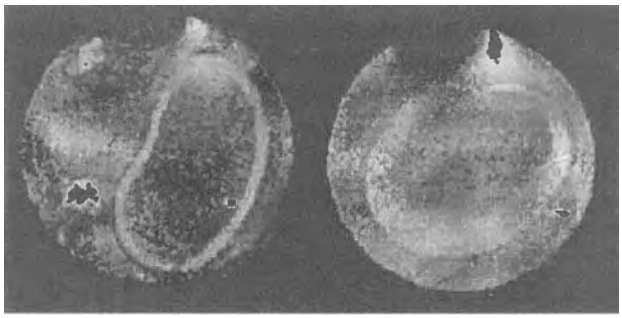


Figure 12. Images of a healthy rat artery (left), and a rat artery with atherosclerosis (right). The images were acquired using multislice two-dimensional spin echo in 25 min with eight averages, and have an in-plane resolution of 13 × 13 μm with a slice thickness of 200 μm.

The time-dependent position of the spin can be expanded in the usual power series,

$$z(t) = z_0 + vt + at^2 + \cdots \tag{28}$$

and the angle of precession is then a moment expansion of the gradient waveform,

$$\Theta_{\text{gradient}} = \gamma\left\{ z_0 \int \frac{\mathrm{d}}{\mathrm{d}z} B_z(t)\,\mathrm{d}t + v \int \frac{\mathrm{d}}{\mathrm{d}z} B_z(t) t\,\mathrm{d}t + a \int \frac{\mathrm{d}}{\mathrm{d}z} B_z(t) t^2\,\mathrm{d}t \right\} \tag{29}$$

Based on this expansion it is straightforward to generate a gradient waveform that will selectively encode the position, velocity, or acceleration of the spin. For example, the gradient echo introduced earlier (see Fig. 6) does not encode the position (this is what is meant by an echo), but it does encode the velocity of the spin. Using gradient modulation methods it is possible to record images with contrast based on spin density, the diffusion coefficient, the velocity, or the acceleration.

1.9 Solid State Imaging

There is a special challenge when exploring solid samples in that the NMR linewidth is

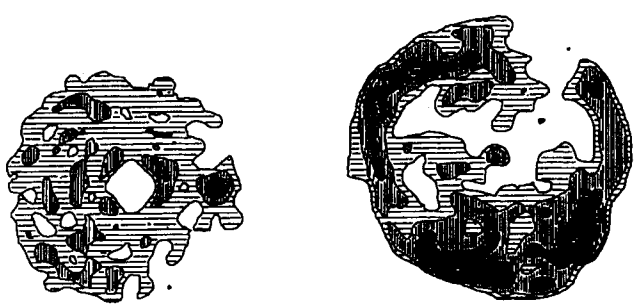


Figure 13. Two-dimensional solid state images of the polybutadiene distribution in a solid, circular disk composed of a polystyrene/polybutadiene blend. Darker intensities correspond to higher butadiene concentrations. The contrast mechanism depended on the strength of the residual homonuclear dipolar coupling, and the diameter of the disk was 4 mm. The in-plane resolution is slightly finer than 50 μm. (Reproduced with permission from D. G. Cory, J. C. de Boer and W. S. Veeman, *Macromolecules* **1989,** *22*, 1618.)

normally quite broad (50 000 Hz), and straightforward applications of the above experiments require strong gradients, and result in a low signal-to-noise ratio. Static gradients of up to 10 000 G cm^{-1} have been achieved and used with back-projection methods, but more powerful methods are often based on artificially narrowing the NMR resonance. Much like a spin echo can be used to reduce the linewidth associated with static variations in the resonance frequency, there are a variety of coherent averaging schemes based on mechanical motions or RF pulse trains that greatly reduce the solid state NMR linewidths. Reductions in the linewidth of factors of 1000 are achievable for plastics and other organic solids, and these schemes have been combined with imaging methods.

Unfortunately, one of the results of most coherent averaging schemes is that the dynamics become highly nonlinear and therefore the experiments themselves become rather complex, in addition to requiring novel instrumentation. Such methods have been reviewed, and the interested reader is referred to these. An example is shown in Fig. 13. The fundamental resolution limit for solid state imaging is not limited by diffusion, as is the case for liquids, but by the sensitivity of the measurement. To date, images have rarely been obtained with finer than 100 μm resolution, but it is thought that the ultimate resolution will be on the order of a few micrometers for small samples.

Acknowledgments

The authors thank the Whitaker Foundation, the National Science Foundation (DMR-9357603), and the National Institutes of Health (RR-00995) for supporting our imaging research, and S. C. acknowledges the receipt of a fellowship from the Korean government. We also thank Dr. S. Gravina for assistance with the experiments, and Dr. R. Kapadia for supplying the samples of rat arteries.

1.10 References

B. Blümich, W. Kuhn, *Magnetic Resonance Microscopy*, VCH, Weinheim **1992**.

P. Blümler, B. Blümich, *NMR—Basic Principles and Progress*, Springer-Verlag, Berlin **1993**.

P. T. Callaghan, *Principles of Nuclear Magnetic Resonance Microscopy*, Oxford University Press, Oxford **1991**.

D. G. Cory, *Annu. Rep. NMR Spectrosc.* **1992**, *24*, 88.

P. Jezzard, J. J. Attard, T. A. Carpenter, L. D. Hall, *Prog. NMR Spectrosc.* **1992**, *23*.

P. Mansfield, E. L. Hahn, *NMR Imaging*, The Royal Society, London **1990**.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

2 Scanning Electron Microscopy with Polarization Analysis (SEMPA)

2.1 Introduction

Scanning electron microscopy with polarization analysis (SEMPA) [1] is a technique that provides high resolution images of magnetic microstructure by measuring the spin polarization of low energy secondary electrons generated in a scanning electron microscope [2–5]. This is possible because the emitted secondary electrons retain the spin polarization present in the material; SEMPA therefore produces a direct image of the direction and magnitude of the magnetization in the region probed by the incident SEM electron beam. SEMPA determines all three components of the spin polarization, and hence of the magnetization. SEMPA records the magnetic and topographic images simultaneously, but independently. Polarization is normalized to the number of emitted electrons, that is, to the intensity, or the quantity measured in a secondary electron SEM topographic image. Thus, SEMPA measurements are intrinsically independent of topography. This feature allows the investigation of the correlation between magnetic and topographic structures. SEMPA can characterize ferromagnetic materials with a sensitivity down to a fraction of an atomic layer and a lateral resolution of 20 nm. The surface sensitivity of SEMPA is particularly advantageous for studies of thin film and surface magnetism [6, 7] but puts requirements on the cleanliness of specimen surfaces. SEMPA also has other advantages common to scanning electron microscopes, such as long working distance, large depth of field, and large range of magnifications. The zoom capability is especially useful for imaging the magnetization distribution in ferromagnets where length scales vary over several orders of magnitude from relatively large ($>10\,\mu m$) magnetic structures such as ferromagnet domains, to intermediate size (200 nm) structures found in Bloch, Néel, asymmetric Bloch or cross-tie domain walls, to the finest structures (<50 nm) found in magnetic singularities such as Bloch lines, Néel caps and magnetic swirls.

To put SEMPA in perspective, it is useful to compare it to other methods of imaging magnetic microstructure, some of which are discussed at length in other chapters of this handbook. Different magnetic imaging methods are distinguished by the quantity measured to obtain magnetic contrast, the resolution, the ease of

Table 1. Comparison of several magnetic imaging techniques

Method	Quantity measured	Resolution [nm]	Information depth [nm]	Reference
SEMPA	magnetization	20	1	[2–7]
magnetooptic Kerr	magnetization	500	10	[16]
MCD	magnetization	300	2	[17]
SPLEEM	magnetization	20	1	[18]
Bitter	field	1000	1000	[8]
MFM	field	10	1000	[15]
TEM Lorentz	field	2	average	[9, 10]
STEM differential phase contrast	field	2	through	[14]
Electron holography	field	2	sample	[12, 13]

interpretation of the measurement, the requirements on sample thickness and surface preparation, the cost, and so on. A summary of this information for the various imaging techniques mentioned below is displayed in Table 1. The values given for the resolution of each technique are estimates of the current state of the art; they should only be taken as a rough guide.

Most methods used for the observation of magnetic microstructure rely on the magnetic fields in and around a ferromagnet to produce magnetic contrast. For example, the oldest method for imaging magnetic microstructure is the Bitter method [8] where fine magnetic particles in solution are placed on the surface of a ferromagnet. The particles agglomerate in the fringe fields at domain walls thereby delineating the magnetic domains; the particles may be observed in an optical microscope or even an SEM. In Lorentz microscopy, the magnetic contrast is derived from the deflection of a focused electron probe as it traverses a ferromagnetic sample [9, 10]. In the transmission electron microscope (TEM), Lorentz microscopy can achieve a high lateral resolution of order 10 nm, but the measurement represents an average over specimen thickness. Only thin samples (<300 nm) are suitable for high spatial resolution studies. Unfortunately, such thin samples may not have a magnetization distribution characteristic of the bulk. Lorentz microscopy in the reflection mode in an SEM has also been demonstrated [11]. It has the advantage that the near surface of bulk specimens can be examined, but the lateral resolution is seldom much better than 1 μm. Electron holography [12, 13] is an electron interferometric method for obtaining absolute values of the magnetic flux in and around thin ferromagnetic samples. It is a high resolution (2 nm) method with contrast derived from the measurement of electron phase shifts that occur in electromagnetic fields. Differential phase contrast microscopy also measures electron phase shifts to give magnetic contrast at high lateral resolution (2 nm) in the scanning TEM [14]. Magnetic force microscopy (MFM) is an imaging technique suitable for thick (bulk) specimens. It achieves contrast through the magnetostatic interaction between a ferromagnetic tip and the fringe fields of the ferromagnet. MFM can be used to locate domain walls with a spatial resolution of

about 10 nm, but it is difficult to extract quantitative information from MFM images [15].

Direct methods for measuring micromagnetic structure rely on contrast mechanisms which reveal the magnetization rather than the magnetic induction. The magneto-optic Kerr effect [16] uses the rotation of the plane of polarization of light upon reflection to map surface magnetization distributions. As an optical method, its spatial resolution is diffraction limited to optical wavelengths, but it has the advantage that a magnetic field can be applied and varied during measurement. Magnetic circular dichroism (MCD) images domains using a photoelectron microscope [17]. Photoelectron images are recorded for circularly polarized incident X-rays. Since the photoelectron yield is proportional to the spin-dependent density of states at the surface and the helicity of the X-rays which selectively excite atomic core levels, images of domain structure can be obtained with elemental specificity. Although the information depth within the magnetic material is about 2 nm, secondaries from a 10 nm carbon overcoat have been found to reflect the underlying magnetic structure [17]. Spin-polarized low energy electron microscopy (SPLEEM) is a very new high resolution (20 nm) method for resolving surface magnetic microstructure [18] which relies on the spin-dependence of the (quasi)elastic scattering cross section for polarized electrons from ferromagnets. A spin polarized electron source is required to modulate the spin of the incident beam. Magneto-optic Kerr, MCD, and SPLEEM are like SEMPA in that they measure quantities directly proportional to the sample magnetization.

2.2 Principle of SEMPA

Scanning electron microscopy with polarization analysis (SEMPA), first demonstrated in the mid-1980s [19–23], is a micromagnetic imaging technique that derives magnetic contrast from the spin polarization of secondary electrons extracted from a ferromagnetic surface. The secondary electron magnetic moments are parallel, and consequently their spins antiparallel, to the magnetization vector at their point of origin in the sample [24].

The SEMPA method is schematically depicted in Fig. 1. As the electron beam is scanned across the sample, the secondary electrons are collected and their polarization analyzed. An electron spin analyzer measures each component of the polarization vector, $\boldsymbol{P}$, separately. For example the x component of polarization is given by

$$P_x = (N_\uparrow - N_\downarrow)/(N_\uparrow + N_\downarrow) \qquad (1)$$

where $N_\uparrow$ ($N_\downarrow$) are the number of electrons detected with spins parallel (antiparallel)

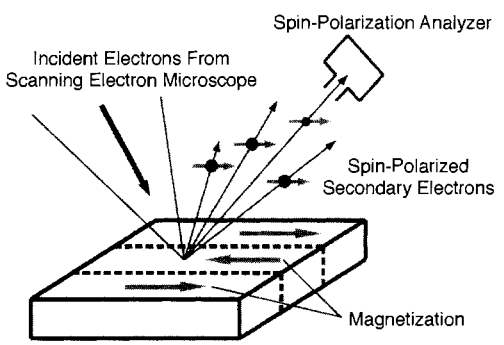


Figure 1. The principle of SEMPA. A scanned beam of electrons incident on the surface of a ferromagnet creates spin-polarized secondary electrons which are subsequently spin-analyzed to yield a high resolution magnetization image.

to the $+x$ direction. The degree of electron spin-polarization varies in the range $-1 \leqslant P \leqslant 1$. Note that P is normalized to the total number of electrons emitted, $(N_\uparrow + N_\downarrow)$. The polarization and intensity are measured simultaneously, but independently. Thus, the magnetic and topographic images are determined separately.

Spin polarized secondary electrons emitted from a ferromagnet reflect the spin part of the magnetization,

$$M = -\mu_B(n_\uparrow - n_\downarrow) \tag{2}$$

Here $n_\uparrow$ $(n_\downarrow)$ are the number of spins per unit volume parallel (antiparallel) to a particular orientation, and μ_B is the Bohr magneton. To the extent that the secondary electron cascade represents a uniform excitation of the valence electrons, the expected secondary electron polarization can be estimated as $P = n_B/n_v$, where n_B is the number of Bohr magnetons per atom and n_v the number of valence electrons per atom. In this way, one estimates a polarization of 0.28, 0.19, and 0.05 for Fe, Co, and Ni, respectively. These agree reasonably well with measurements of secondary electrons in the 10 to 20 eV range [25, 26]. At lower energies, spin dependent scattering [27] actually increases the polarization, improving the contrast of SEMPA measurements. SEMPA provides vector magnetization maps for conducting or semiconducting bulk specimens, and thin films and monolayer films where specimen charging is not a problem. In some cases charging can be avoided and the magnetization enhanced by evaporating a 1 nm thick film of Fe on the specimen; this has allowed the imaging of insulating Fe garnets [28]. The magnetization along the measurement direction is proportional, but oppositely directed, to the electron polarization along that direction. In practice, the constant of proportionality is not precisely known; the detailed scattering dynamics for the production of polarized secondary electrons is dependent upon the surface band structure, which varies from material to material. Measurements using SEMPA reveal the spatial dependence of the relative value of the surface magnetization distribution rather than the absolute size of the surface moments.

Important features of SEMPA include its high spatial resolution and its surface sensitivity. The spatial resolution of SEMPA is primarily determined by the incident beam diameter focused on the sample surface. Even though the profile of the energy deposited in bulk samples expands greatly in the bulk due to multiple scattering [11], the escape depth of polarized secondary electrons is on the order of nanometers. The distance over which the secondary electron spin polarization is exponentially reduced by 1/e is about 0.5 nm for a transition metal like Cr [29], and about 1.5 nm for a noble metal like Ag [30]. Although SEMPA measures only the near-surface micromagnetic structure, the underlying and bulk magnetic structure can be determined by solution of the micromagnetic equations using the surface magnetization measurements as boundary conditions [31]. The surface sensitivity of SEMPA is advantageous for studies of surface and thin film magnetism but can be a limitation in the sense that sample surfaces must be clean. Thick oxides or hydrocarbon layers will diminish the polarization and hence the magnetic contrast.

2.3 Instrumentation

The electron probe forming column, transport optics, and spin-polarization detectors comprise the essential electron optical components of the SEMPA system. A schematic of a SEMPA instrument is shown in Fig. 2. Because of the surface sensitivity of SEMPA, the specimen should be cleaned and maintained in ultrahigh vacuum ($P < 5 \times 10^{-8}$ Pa). Conventional surface science preparation and analysis tools including an ion-beam sputtering gun, an electron beam evaporator, an Auger electron spectrometer and a reflection high energy electron diffraction screen greatly facilitate the preparation and characterization of the sample surface. The SEMPA system may be equipped with a single spin detector [4–7], or multiple spin detectors [2, 3] as shown in Fig. 2. Two detectors are used for the acquisition of all three orthogonal components of the vector polarization (magnetization) signal.

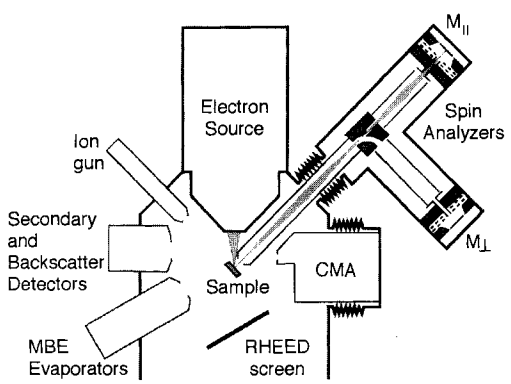


Figure 2. Schematic of a SEMPA apparatus. The electron source, polarization detectors, cylindrical mirror analyzer (CMA), and reflection high energy electron diffraction (RHEED) screen are shown in their actual relative positions; the rest of the instruments are not. The CMA and polarization analyzers are retractable.

The SEM probe forming column, the transport optics, the polarization analyzers, the electronics and signal processing will be considered in turn. A general description will be given of the generic components of SEMPA; we will use our apparatus as an example for detailed analysis.

2.3.1 Scanning Electron Microscopy Probe Forming Column

An SEM beam of 10 keV is a reasonable compromise among the constraints of secondary electron yield, spatial resolution, and beam stability in the secondary electron extraction field. The beam energy must be high enough to reduce the deleterious effects of electron lens aberrations, yet low enough to sustain reasonable secondary electron yields from the sample. Submicrometer beam diameters can be obtained for electron energies above 5 keV while the secondary electron yield, for example from Al, falls from 0.40 at an incident beam energy of 5 keV to 0.05 at 50 keV [11]. The incident electron beam must also be energetic enough such that the extraction optics which transport the polarized secondary electrons to the spin detector do not severely aberrate the focused spot on the sample. Extraction optics typically have fields on the order of 100 V/mm to achieve adequate collection efficiency. A 10 keV beam suffers minimal distortion in such an extraction field.

Two essential components of the electron optical column, the electron source and the probe forming objective lens, can be optimized for SEMPA. For reasonable

SEMPA acquisition times, electron sources must provide a 10 keV incident beam with a current of at least 1 nA to the specimen. It is this constraint that determines the SEMPA spatial resolution rather than the ultimate resolution of the column when used as a standard SEM. The selection of an electron source rests on the spatial resolution required for a specific micromagnetic measurement. Lanthanum hexaboride (LaB_6) [2, 3], cold field emission cathodes [4–7], and thermal field emitters [32] have all been employed in SEMPA. The highest resolution, approximately 20 nm, has been achieved with field emitters. Thermal field emitters [33, 34] have somewhat larger source size than cold field emitters, but have greater stability (current variations <1%), high emission currents, and moderate energy width, making them well suited for use in high resolution SEMPA systems.

In SEM columns, the spherical aberration of the probe forming objective lens increases rapidly with increasing working distance, the distance between the lens exit pole face and the sample. For high resolution, one wants a short working distance. On the other hand, longer working distances are desirable to obtain a region at the sample surface free (<80 A/m) from the depolarizing effects of the stray magnetic field of the objective lens. Working distances between 5 and 15 mm provide a satisfactory trade off.

2.3.2 Transport Optics

The purpose of the transport optics is to efficiently collect and transfer the spin-polarized secondary electrons from the specimen surface to the spin-polarization detectors without introducing instrumental asymmetries. Instrumental asymmetries are systematic errors which may be accounted for in a variety of ways. To reduce the effects of chromatic aberrations on the transported beam, the secondary electrons are first accelerated in a potential greater than about 500 V. In order to achieve the highest possible efficiency, the transport energy window of the electron optics should be about 8 eV wide and centered at 4 eV. The optical properties of low energy electron lenses used for transport can be computed from the numerical solution of Laplace's equation and subsequent ray tracing of the charged particle trajectories through the fields. An invaluable compendium of electron optical properties of common electron lens configuration has been compiled by Harting and Read [35]. At low magnification, the motion of the incident SEM beam on the specimen is translated into motion of the beam on the spin analyzer target leading to an instrumental asymmetry; a dynamic beam descanning scheme can be employed to remove scan related asymmetries [3]. The transport optics can be further optimized to reduce instrumental asymmetries and compensate for variations in the position of the beam [36].

2.3.3 Electron Spin Polarization Analyzers

Ideally a spin-polarimeter suitable for SEMPA should be efficient, small in size and compatible with the UHV ambient required for sample preparation. There has been considerable progress in reducing

the size and increasing the efficiency of electron spin polarimeters, yet spin detectors remain quite inefficient [37]. The polarization of a beam of N_0 electrons is $P_0 \pm \delta P$, where the uncertainty is $\delta P = 1/(N_0 F)^{1/2}$. The figure of merit, F, for a spin polarization analyzer, rarely is much greater than 10^{-4} even for the highest performance spin detectors. Thus, a polarization measurement with a relative uncertainty, $\delta P/P_0 = 1/(P_0^2 N_0 F)^{1/2}$, equivalent to the relative uncertainty in an intensity measurement, $\delta N/N_0 = 1/(N_0)^{1/2}$, can take over 10^4 times as long as the intensity measurement solely due to the inefficiency of the electron polarimeter.

Most spin polarimeters rely on a spin–orbit interaction for spin sensitivity. When an electron scatters from a central potential $V(\boldsymbol{r})$, the interaction of the electron spin $\boldsymbol{s}$ with its own orbital angular momentum $\boldsymbol{L}$ [38] has the effect of making the cross section larger or smaller for electrons with spin parallel or antiparallel to $\boldsymbol{n}$, the unit vector normal to the scattering plane. The scattering plane is defined by the incident electron wave-vector $\boldsymbol{k}_i$ and scattered electron wavevector $\boldsymbol{k}_f$ such that $\boldsymbol{n} = (\boldsymbol{k}_i \times \boldsymbol{k}_f)/|\boldsymbol{k}_i \times \boldsymbol{k}_f|$. The cross section for the spin-dependent scattering can be written [38] as

$$\sigma(\Theta) = I(\Theta)[1 + S(\Theta)\boldsymbol{P} \cdot \boldsymbol{n}] \qquad (3)$$

where $I(\Theta)$ is the angular distribution of back scattered current in the detector and $S(\Theta)$ is the Sherman function for the detector scattering material at the scattering angle, Θ. The Sherman function is a measure of the strength of the spin-dependent scattering in the detector [39]. Typical values for S are $|S| < 0.3$. The polarization of the beam is determined from a spatial asymmetry A between the number of electrons scattered to the left, N_L, and to the right, N_R, relative to the incident beam direction. The measured scattering asymmetry, A, is

$$A = (N_L - N_R)/(N_L + N_R) = PS \qquad (4)$$

Differences in the left/right scattering can also arise from instrumental asymmetries and cause systematic errors that contribute to the uncertainty in the polarization measurement. These instrumental asymmetries result from: (i) unequal gains in the left and right channels of the signal processing electronics; (ii) unequal sensitivities of the electron multipliers; and (iii) mechanical imperfections which result in a detector geometry that is not symmetric.

As an example of a scattering type spin analyzer, we describe the low energy diffuse scattering (LEDS) detector [36, 40] used in our work. A schematic of this analyzer is shown in Fig. 3. The analyzer is quite compact since it operates at 150 eV; in our design it is about 10 cm

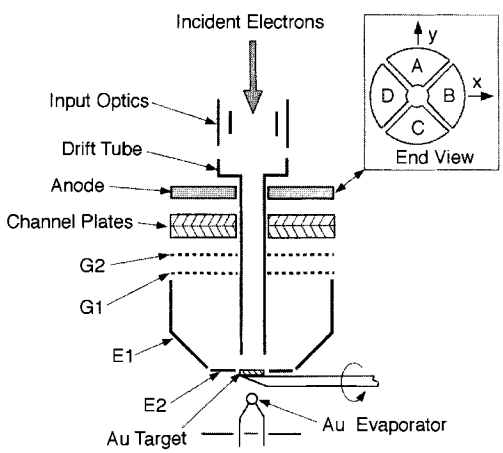


Figure 3. Schematic of the low energy diffuse scattering spin polarization analyzer. The divided anode assembly is shown in the inset as viewed from the Au target.

long. It employs an evaporated polycrystalline Au target. The efficiency of this spin analyzer is increased by collecting the scattered electrons over large solid angles. The large ratio of electrons collected to the number incident compensates for the moderate Sherman function of about 0.1, and leads to a high figure of merit, 2×10^{-4} [36]. The electrodes E1 and E2 in Fig. 3 focus the electrons that are diffusely backscattered from the Au target such that their trajectories are nearly normal to the retarding grids G1 and G2. The energy selectivity of the retarding grids enhances the Sherman function while the collection efficiency is increased by electrodes E1 and E2. The anode, which is also shown in the inset of Fig. 3, is divided into quadrants. Two orthogonal components of the polarization vector transverse to the electron beam direction (z) may be measured simultaneously with this detector as

$$P_x = 1/S(N_C - N_A)/(N_C + N_A) \qquad (5a)$$

$$P_y = 1/S(N_B - N_D)/(N_B + N_D) \qquad (5b)$$

where N_i is the number of electrons counted by each quadrant (i = A, B, C, D).

No single electron spin analyzer has all the features one might desire for highest performance in a SEMPA application. The traditional Mott analyzer which utilizes the asymmetry of the spin-dependent high energy (100–200 keV) electrons [38, 41, 42] has a Sherman function S that is larger than that of the LEDS detector and a comparable figure of merit. The high energy operation makes it less susceptible to apparatus asymmetries than low energy spin analyzers which therefore require more care in design of the transport optics. However, operation at the required high voltage leads to large detector sizes making the Mott analyzer challenging to integrate with the SEM. Nevertheless, such analyzers have been used quite successfully for SEMPA [4, 7]. A low energy electron diffraction (LEED) electron-spin polarization analyzer [43, 44] has also been used very successfully for SEMPA [5, 6]. The collimating properties of diffraction by a single crystal, usually W(1 0 0) at about 100 eV, increase the efficiency of this spin analyzer leading to a relatively compact analyzer with a competitive figure of merit.

2.3.4 Electronics and Signal Processing

The electron signal is measured with surface barrier Si detectors, channeltrons, or stacked microchannel plates with a segmented anode, respectively, in the Mott, LEED, and LEDS spin analyzers. Each pair of detectors determines a component of the spin polarization vector transverse to the beam. For pulse counting, each signal channel consists of a preamplifier, amplifier/discriminator, and a scaler that is read by the computer. The signal processing electronics for the LEDS spin analyzer have been realized in both the pulse counting and analog modes; we describe aspects of each of these methods below.

When the electron probe beam is focused to very high spatial resolution in the SEM column, the beam current is reduced and pulse counting in the polarimeter is necessary. The short pulse widths (above 1 ns) in stacked microchannel plates facilitate high speed counting.

Dark currents are typically less than 1 count $cm^{-2} s^{-1}$. In pulse counting, the quadrant anode structure must be designed to minimize cross talk between adjacent channels. Otherwise, pulses from adjacent channels will appear at reduced amplitude making pulse discrimination difficult. Reduction of inter-anode capacitance and capacitive coupling from each anode to common surrounding surfaces is effective in reducing this problem. Fast (20 ns) charge sensitive preamps can also be used. These are less affected by crosstalk than voltage preamps because they average over the characteristic ringing signal of capacitively coupled cross talk.

Fast, low resolution magnetization imaging with high incident current is very useful to survey a sample. At high incident beam currents, analog signal processing becomes necessary since the microchannel plates are count rate limited [45, 46]. With separate direct-coupled outputs from the anode quadrants, it is straightforward to switch over to an analog measurement of the current to the quadrants. When analog detection is used at high incident beam currents, the channel plate bias voltage is decreased to maintain linear gain. Each anode pair is connected to signal processing electronics including matched current-to-voltage converters, and sum and difference amplifiers. (Alternatively, the sum and difference can be performed later in the computer). The sum and difference signals are converted to pulse trains by separate voltage to frequency converters. Optocouplers provide the isolation necessary for the input stages to operate at the microchannel plate anode voltage and the signals are counted with a conventional scaler and timer system. Since the difference signal may change sign, an offset voltage is applied to that voltage-to-frequency converter to prevent zero crossing and minimize digitization errors [3].

2.4 System Performance

The performance of a SEMPA system can be analyzed by examining the efficiency of the entire production, collection and processing chain. Although some of the considerations in the analysis are generally applicable to any SEMPA system, in order to provide specific numbers we will give parameters for our SEMPA system with the LEDS spin analyzer [2, 3]. The production efficiency of secondary electrons by a 10 keV electron beam at the surface of a ferromagnetic specimen tilted by 45°, is roughly 0.45 [11]. Only 37% of the secondary electrons produced at the sample are collected since the extraction optics only collect a narrow secondary electron energy window, $4.0\,eV \pm 4.0\,eV$. The efficiency of the transport optics between the extraction aperture and the spin analyzer may be as high as 1.00, but for normal operation the transport efficiency is closer to 0.88. The scattering efficiency, or ratio of the current incident on the detector channel plate input to that incident on the Au target, is 0.04 for nominal operating conditions [36, 40]. The channel plate itself has a finite collection efficiency of about 0.85 [46] due to final cell size. The product of all of these factors is the collection efficiency of the system, $\epsilon = 0.005$. For 1 nA incident beam current, only 4 pA ($0.004\,I_0$) of secondary electrons will be detected in the electron polarimeter, or approximately 1 pA to each quadrant.

Signal levels and integration times required to reach selected signal-to-noise ratios can also be estimated [3]. The simplest case to consider is the image of two adjacent domains with oppositely directed magnetization. Assume that the sample is oriented along a single detector direction such that the measured component of the polarization will be $+P$ in one domain and $-P$ in the other. The total change in that polarization component between the two domains (i.e., the signal) is $2P$. For a polarization measurement limited by counting statistics [38], one standard deviation statistical error in the polarization, $P = (1/S)(N_C - N_A)/(N_C + N_A)$, is given by $\delta P = (1/(N_C + N_A)S^2)^{1/2}$. The particle number reaching any pair of detector quadrants $(N_A + N_C)$ in a time interval τ is $\epsilon(I_0/2e)\tau$. The signal-to-noise ratio is

$$SNR \equiv 2P/\delta P = 2PS(\epsilon I_0 \tau/2e)^{1/2} \qquad (6)$$

The dwell time required for each pixel in the image as a function of the desired signal-to-noise ratio and the experimental parameters is

$$\tau = (SNR)^2 e/(2P^2 S^2 \epsilon I_0) \qquad (7)$$

The upper limit on the count rate will be set by the channel plate response. Assuming that the incident electron beam current in the electron microscope column is $I_0 = 1\,\text{nA}$ and $S = 0.11$, the dwell time per pixel for various signal-to-noise ratios, and polarizations is given in Table 2. The elements in the table must be multiplied by the number of pixels in an image for the total data acquisition time. Thus, it takes about 54 s to acquire a 256×256 pixel image with a signal-to-noise ratio of 5 and a mean polarization of 0.20. For analog signal detection, it is possible to reduce the noise introduced by the analog amplifier well below the shot noise of the incident beam for incident currents $\geqslant 1\,\text{nA}$. Hence the dwell times given in Table 2 also apply for analog signal acquisition.

Table 2. Pixel dwell time, τ (ms), as a function of the *SNR* and polarization for SEM beam current, $I_0 = 1\,\text{nA}$

P	$SNR = 2$	$SNR = 3$	$SNR = 5$	$SNR = 10$
0.01	53.185	119.666	332.407	1329.626
0.10	0.532	1.197	3.324	13.296
0.20	0.133	0.299	0.831	3.324
0.40	0.033	0.075	0.208	0.831

2.5 Data Processing

Conventional image processing methods, such as filtering and contrast enhancement, can be used for processing SEMPA images. There are, however, some image processing steps that are unique to SEMPA since the contrast is derived from a vector magnetization and the spin detector sensitivity results from a scattering asymmetry. For SEMPA, common image processing steps include the subtraction of a zero offset and a background asymmetry. To do this, use is made of the fact that the magnitude of the magnetization $|M|$ is constant. Consider the common case where the magnetization vector lies entirely in-plane. (This is expected for all but materials with particularly large magnetic anisotropy perpendicular to the surface.) In this case, it is possible to subtract a background and remove zero offsets by requiring that the in-plane magnetization, $(M_x^2 + M_y^2)^{1/2}$, have constant magnitude. In general, the

background subtracted may be nonlinear and fit with a polynomial. In special cases, the specimen topography can cause trajectories which couple with instrumental asymmetries to produce artifacts in the polarization measurements. In such cases the polarization detector's gold target can be replaced by a low atomic number target, such as graphite, for which there is no spin-dependent scattering asymmetry. The image acquired with the graphite target is then subtracted from that acquired with the Au target to remove instrumental asymmetries.

There are two basic formats to represent SEMPA data. The first uses the projection of the magnetization on orthogonal axes (i.e., M_x and M_y) and uses a gray map encoding scheme where white (black) represents the maximum value of the magnetization in the positive (negative) directions. The second format uses the magnitude $|M|$ and the angle ϑ of the magnetization vector projected onto some plane. Whether it is easier to identify a surface magnetic domain structure in M_x and M_y images or in $|M|$ and ϑ images depends largely upon the surface magnetic microstructure. The magnitude of the magnetization is determined as

$$|M| = [M_x^2 + M_y^2 + M_z^2]^{1/2} \tag{8}$$

and the direction, with respect to the positive x-axis of the in-plane magnetization, is (in the absence of any out-of-plane components)

$$\vartheta = \tan^{-1}(M_y/M_x) \tag{9}$$

The map of the angle ϑ can be displayed using color where the direction is read from an accompanying color wheel. Alternatively, it is sometimes helpful to visualize the magnetization pattern by using small arrows to create a vector map.

2.6 Examples

2.6.1 Iron Single Crystals

The large magnetic moment per atom of Fe leads to a large intrinsic secondary electron polarization which makes Fe a favorable specimen to use for demonstrating SEMPA features. Figure 4 shows SEMPA images [2] of the (1 0 0) surface of a high quality Fe single crystal whisker. In the image labeled I, one observes the flat featureless upper surface of the whisker running vertically, centered in the frame. The image of the x component of the magnetization, M_x, shows a diamond-shaped domain with magnetization pointing to the right. The domain pattern is

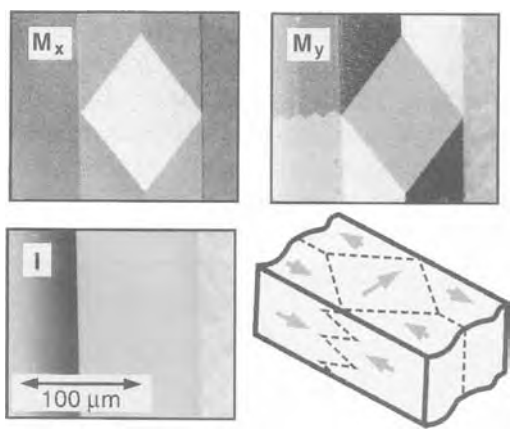


Figure 4. SEMPA measurements of the topography, *I*, and images of two components of the magnetization, M_x and M_y, from an Fe whisker. The depth of focus is demonstrated by domains clearly visible on the top and side of this slightly tilted sample of rectangular cross section.

shown schematically in the line drawing. The region to the right of the diamond in the figure is the nonmagnetic sample holder; to the left is the side of the crystal. The image of the vertical component of the magnetization, M_y, contains domains with magnetization pointing in the $+y$ direction (white) and in the $-y$ direction (black). The zig–zag domain wall visible in this image actually runs down the side of the whisker and is visible because the sample is slightly tilted. This large depth of focus is characteristic of SEMPA since a scanning electron microscope is used as the probe.

2.6.2 CoPt Magnetooptic Recording Media

In magnetooptic recording, information is stored by using a focused laser beam to read and write the bits. A bit is written when the laser locally heats the media in an applied magnetic field. The result is seen in Fig. 5 where the white dots correspond to magnetization out of plane, M_z, in the $+z$ direction contrasted against a background previously magnetized in the $-z$ direction. The corresponding intensity image shows the nonuniform topography of the Co-Pt multilayer sample. In use, the bits are read by sensing the rotation of polarization of reflected light from the surface. The circular bits observed in Fig. 5 are about 1.4 μm in diameter. A large perpendicular magnetic anisotropy is necessary to overcome the increased magnetostatic energy of out-of-plane magnetization. Domains with out-of-plane magnetization have also been observed on Co(0 0 0 1) single crystals [47] and on TbFeCo magnetooptic storage media [48].

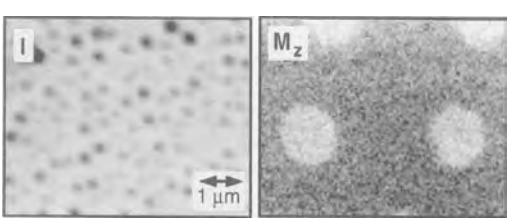


Figure 5. The topography, I, and perpendicular magnetization, M_z, are imaged from an CoPt magnetooptic recording medium. The round bits are approximately 1.4 μm in diameter.

2.6.3 Exchange Coupling of Magnetic Layers

Two ferromagnetic layers separated by a nonferromagnetic layer may be exchange coupled such that the magnetic moments in the two ferromagnetic layers are parallel (ferromagnetic exchange coupling) or antiparallel (antiferromagnetic exchange coupling) depending on the spacer layer material and its thickness. An example is two Fe layers separated by a Cr layer to form a sandwich structure Fe/Cr/Fe(1 0 0). SEMPA is particularly well suited to determine the period (or periods) of oscillation of the exchange coupling between the magnetic layers as a function of spacer layer thickness [49, 50]. For example, consider the Fe/Cr/Fe(1 0 0) sandwich structure shown schematically in Fig. 6. A varying thickness Cr 'wedge' is deposited on the Fe(1 0 0) whisker substrate. This is covered with an Fe film approximately 10 layers thick. As shown in the schematic, the Fe layers are ferromagnetically coupled for small Cr thickness, and the sign of the coupling oscillates as the Cr thickness increases.

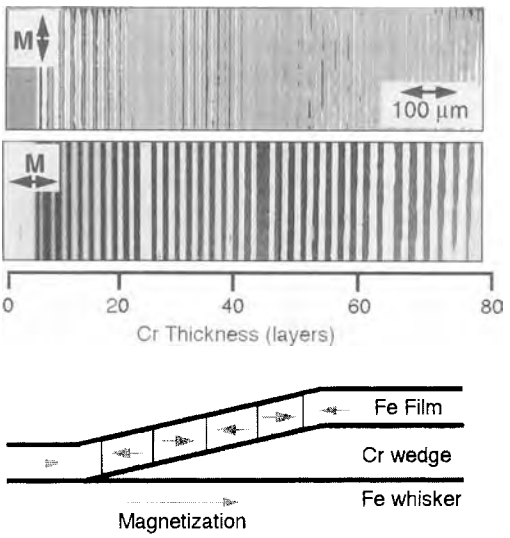


Figure 6. The magnetization images, M_x and M_y, of the top layer of a Fe/Cr/Fe(1 0 0) sandwich structure, shown schematically at the bottom, provide a precise measure of the oscillation of the sign of the magnetic exchange coupling as the thickness of the Cr spacer layer increases.

The SEMPA image of the magnetization in the direction of the wedge shows many changes in the magnetization as the Cr increases in thickness to nearly 80 layers (1 layer = 0.14 nm) over the length of the wedge shown in the image, about 0.8 mm. Such an image allows a very precise determination of the period of the exchange coupling. At Cr thicknesses where there is a reversal in the magnetic coupling, one observes a component of the magnetization transverse to the wedge direction as seen in the upper magnetization image. This provided early evidence for a different kind of coupling known as biquadratic exchange coupling.

Several features of SEMPA were successfully exploited in these studies. The high spatial resolution of the SEM allows the use of a small, nearly perfect Fe single crystal whisker substrate. The ultrahigh vacuum allows deposition of a Cr wedge in situ. The clear advantage of the wedge structure is that it allows measurements at many different thicknesses with a reproducibility that could not be obtained by producing multiple films of uniform thickness. With a reflection high energy diffraction (RHEED) screen below the sample stage, and using the SEM column as a RHEED gun, it is possible to make spatially resolved RHEED measurements along the Cr wedge to determine the thickness of the interlayer material with single atomic layer resolution. The surface sensitivity of SEMPA allows the observation of the changes in the magnetization direction of the Fe overlayer without any interference from the lower Fe layer. Finally, since the SEMPA system used for these measurements was part of a scanning Auger microprobe, Auger spectroscopy could be utilized to monitor cleanliness of the specimen at each stage of its preparation.

2.6.4 Magnetic Singularities in Fe–SiO_2 Films

A high resolution field emission SEMPA was used to image magnetic singularities in granular Fe–SiO_2 films above the percolation threshold [51]. These are nanocomposite materials with highly isotropic magnetic properties [52]. Figure 7 shows a segment of a 180° domain wall with two cross ties and a dramatic ripple pattern. Fig. 7a depicts the x component of the magnetization. This image clearly shows the fine structure of the ripple pattern; the cross ties appear as diamond shaped regions elongated in the direction

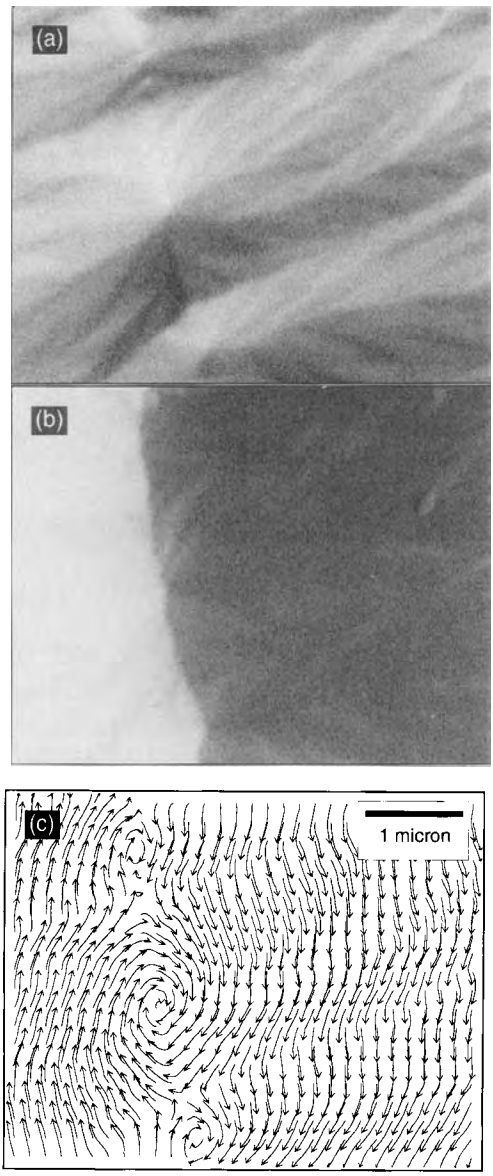


Figure 7. SEMPA images of the in-plane magnetization in a Fe–SiO_2 film near a domain wall: (a) x-component image; (b) y-component image; (c) vector map representation.

orthogonal to the wall. In the y component image, Fig. 7b, the wall itself is readily apparent. The vector map representation given in Fig. 7c shows the coarse structure of the ripple and the vortices that occur between cross ties.

Because SEMPA is only sensitive to the magnetization at the surface, it is ideally suited to an investigation of surface features such as the singularities in Fig. 7. Line scans taken across the large central vortex indicate that the in-plane component of the magnetization decreases within the vortex showing that there must be a perpendicular component to the magnetization. The resolution of these images was determined to be 20 nm by analysis of line scans across several surface features.

Acknowledgements

The iron whiskers were grown at Simon Fraser University under an operating grant from the National Science and Engineering Research Council of Canada. We thank all participating members of the Electron Physics Group for numerous contributions. This work was supported in part by the Office of Naval Research.

2.7 References

[1] This technique is sometimes referred to as spin polarized SEM. We prefer the name SEMPA because the SEM beam is not spin polarized.
[2] J. Unguris, M. R. Scheinfein, R. J. Celotta, D. T. Pierce in *Chemistry and Physics of Solid Surfaces VII* (Eds.: R. Vaneslow, R. Howe), Springer Verlag, Germany **1990,** p. 239.
[3] M. R. Scheinfein, J. Unguris, M. H. Kelley, D. T. Pierce, R. J. Celotta, *Rev. Sci. Instrum.* **1990,** *61*, 2501.
[4] H. Matsuyama, K. Koike, *Rev. Sci. Instrum.* **1991,** *62*, 970; *J. Electron. Microsc.* **1994,** *43*, 157.
[5] H. P. Oepen, J. Kirschner, *Scanning Microsc.* **1991,** *5*, 1.
[6] H. P. Oepen, *J. Magn. Magn. Mat.* **1991,** *93*, 116.
[7] R. Allenspach, *J. Magn. Magn. Mat.* **1994,** *129*, 160.
[8] F. Bitter, *Phys. Rev.* **1931,** *38*, 1903.

[9] K. Tsuno, *Rev. Solid State Science* **1988,** *2*, 623.
[10] J. P. Jacubovics, *Electron Microscopy in Materials Science Part IV* (Eds.: E. Ruedle, U. Valdre), Commission of European Communities, Brussels **1973,** p. 1305.
[11] D. E. Newbury, D. E. Joy, P. Echlin, C. E. Fiori, J. I. Goldstein, *Advanced Scanning Electron Microscopy and X-ray Microanalysis,* Plenum Press, New York **1986.**
[12] A. Tonomura, *Rev. Mod. Phys.* **1987,** *59*, 639.
[13] M. Mankos, M. R. Scheinfein, J. M. Cowley, *J. Appl. Phys.* **1994,** *75*, 7418.
[14] J. N. Chapman, *J. Phys.* **1984,** *D17*, 623; J. N. Chapman, R. Ploessl, D. M. Donnet, *Ultramicroscopy* **1992,** *47*, 331; G. R. Morrison, H. Gong, J. N. Chapman, V. Hrnciar, *J. Appl. Phys.* **1988,** *64*, 1338.
[15] P. Grutter, H. J. Mamin, D. Rugar, *Springer Ser. Surf. Sci.* **1992,** *28*, 151.
[16] J. Kranz, A. Z. Hubert, *Z. Angew. Phys.* **1963,** *15*, 220.
[17] J. Stohr, Y. Wu, B. D. Hermsmeier, M. G. Samant, G. R. Harp, S. Koranda, D. Dunham, B. P. Tonner, *Science* **1993,** *259*, 658.
[18] K. Grzelakowski, T. Duden, E. Bauer, H. Poppa, S. Chiang, *IEEE Trans. Magn.* **1994,** *30*, 4500.
[19] R. J. Celotta, D. T. Pierce, *Microbeam Analysis–1982* (Ed.: K. F. J. Heinrich), San Francisco Press, San Francisco **1982**, p. 469.
[20] K. Koike, K. Hayakawa, *Jpn. J. Appl. Phys.* **1984,** *23*, L187.
[21] J. Unguris, G. G. Hembree, R. J. Celotta, D. T. Pierce, *J. Microscopy* **1985,** *139*, RP 1.
[22] K. Koike, H. Matsuyama, K. Hayakawa, *Scanning Micros. Supp.* **1987,** *1*, 241.
[23] G. G. Hembree, J. Unguris, R. J. Celotta, D. T. Pierce, *Scanning Microsc. Supp.* **1987,** *1*, 229.
[24] J. Unguris, D. T. Pierce, A. Galejs, R. J. Celotta, *Phys. Rev. Lett.* **1982,** *49*, 72.
[25] H. Hopster, R. Raue, E. Kisker, G. Guntherodt, M. Campagna, *Phys Rev. Lett.* **1983,** *50*, 70.
[26] E. Kisker, W. Gudat, K. Shroder, *Solid State. Commun.* **1982,** *44*, 591.
[27] D. R. Penn, S. P. Apell, S. M. Girvin, *Phys. Rev. Lett.* **1985,** *55*, 518; *Phys. Rev.* **1985,** *32*, 7753.
[28] J. Unguris, M. W. Hart, R. J. Celotta, D. T. Pierce, Proc. 49th EMSA (Ed.: G. W. Bailey), San Francisco Press, San Francisco **1991**, p. 764.
[29] J. Unguris, R. J. Celotta, D. T. Pierce, *Phys. Rev. Lett.* **1992,** *69*, 1125.
[30] J. Unguris, R. J. Celotta, D. T. Pierce, *J. Mag. Mag. Mat.* **1993,** *127*, 205.
[31] M. R. Scheinfein, J. Unguris, J. L. Blue, K. J. Coakley, D. T. Pierce, R. J. Celotta, P. J. Ryan, *Phys. Rev.* **1991,** *B43*, 3395.
[32] A. Gavrin, M. H. Kelley, J. Q. Xiao, C. L. Chien, *Appl. Phys. Lett.* **1995,** *66*, 1683.
[33] D. W. Tuggle, J. Z. Li, L. W. Swanson, *J. Microsc.* **1985,** *140*, 293.
[34] D. W. Tuggle, L. W. Swanson, *J. Vac. Sci. Technol.* **1985,** *B3*, 193.
[35] E. Harting, F. H. Read, *Electrostatic Lenses*, Elsevier, Amsterdam, **1976.**
[36] M.R. Scheinfein, D. T. Pierce, J. Unguris, J. J. McClellend, R. J. Celotta, *Rev. Sci. Instr.* **1989,** *60*, 1.
[37] D. T. Pierce, R. J. Celotta, M. H. Kelley, J. Unguris, *Nuc. Inst. Meth.* **1988**, A266, 550.
[38] J. Kessler, *Polarized Electrons, 2nd ed.*, Springer-Verlag, Berlin **1985.**
[39] M. Fink, A. C. Yates, *At. Data* **1970,** *1*, 385; M. Fink, J. Ingram, *At. Data* **1972,** *4*, 129.
[40] J. Unguris, D. T. Pierce, R. J. Celotta, *Rev. Sci. Instrum.* **1986,** *57*, 1314.
[41] E. Kisker, R. Clauberg, W. Gudat, *Rev. Sci. Instrum.* **1982,** *53*, 1137.
[42] G. D. Fletcher, T. J. Gay, M. S. Lubell, *Phys. Rev.* **1986,** *A34*, 911.
[43] J. Kirschner, *Polarized Electrons at Surfaces*, Springer-Verlag, Berlin, **1985.**
[44] J. Sawler, D. Venus, *Rev. Sci. Instrum.* **1991,** *62*, 2409.
[45] J. L. Wiza, *Nucl. Instrum. Meth.* **1979,** *162*, 578.
[46] E. H. Eberhardt, *IEEE Trans. Nucl. Sci.* **1981,** *NS 28*, 712.
[47] J. Unguris, M. R. Scheinfein, R. J. Celotta, D. T. Pierce, *Appl. Phys. Lett.* **1989,** *55*, 2553.
[48] M. Aeschlimann, M. Scheinfein, J. Unguris, F. J. A. M. Greidanus, S. Klahn, *J. Appl. Phys.* **1990,** *68*, 4710.
[49] J. Unguris, R. J. Celotta, D. T. Pierce, *Phys. Rev. Lett.* **1991,** *67*, 140.
[50] D. T. Pierce, J. Unguris, R. J. Celotta in *Ultrathin Magnetic Structure II* (Eds.: B. Heinrich, J. A. C. Bland), Springer-Verlag, Berlin **1994**, p. 117.
[51] A. Gavrin, J. Q. Xiao, C. L. Chien, to be published.
[52] C. L. Chien, *J. Appl. Phys.* **1991,** *69*, 5267.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

3 Spin-Polarized Low-Energy Electron Microscopy

3.1 Introduction

Spin-polarized low-energy electron microscopy (SPLEEM) is a method for imaging the magnetic microstructure of surfaces and thin films with slow specularly reflected electrons. It is based on the fact that electron scattering is spin-dependent via the spin–spin and spin–orbit interactions between the incident electron and the specimen. In specular reflection only the spin–spin interaction occurs. If the specimen has regions with preferred spin alignment at length scales at or above the resolution limit of low-energy electron microscopy (LEEM), then these regions can be imaged via the contribution of the spin–spin interaction to the total scattering potential. Thus, magnetic and structural information is obtained simultaneously. SPLEEM is easily combined with low-energy electron diffraction (LEED), mirror electron microscopy (MEM) and the various types of emission microscopy (photo electron, secondary electron or thermionic electron emission microscopy). Lateral and depth resolution, information depth, and field of view are comparable to that of LEEM (see Chap. 4, Sec. 1.6 of this Volume). Therefore, SPLEEM is an excellent method for the study of the correlation between magnetic structure, microstructure, and crystal structure.

3.2 Theoretical Foundations

Polarized electrons [1] have been used for some time in the study of the structure and magnetism of surfaces by spin-polarized low-energy electron diffraction (SPLEED) [2–5]. SPLEED is a laterally averaging method, but if the sample is used as a cathode in a cathode lens electron microscope the diffracted electrons may be used for imaging the surface in the same manner as in a standard LEEM instrument. The difference between SPLEEM and LEEM results from the fact that the incident beam is spin polarized in SPLEEM. The exchange interaction $V_{ex} = \sum_i J(\boldsymbol{r} - \boldsymbol{r}_i)\boldsymbol{s} \cdot \boldsymbol{s}_i$, where $\boldsymbol{s}$, $\boldsymbol{s}_i$ are the spins of the incident and target electron, $\boldsymbol{r}$, $\boldsymbol{r}_i$ their positions, and $J(\boldsymbol{r} - \boldsymbol{r}_i)$ the exchange coupling strength, does not average to zero in regions with preferred spin alignment $\boldsymbol{s}_i$ because of the polarization $\boldsymbol{P}$ of the incident beam ($\boldsymbol{s}$). If $\boldsymbol{M}$ is the

magnetization resulting from the preferred alignment of the target spins s_i, then $V_{ex} \sim \boldsymbol{P} \cdot \boldsymbol{M}$, which causes a $\boldsymbol{P} \cdot \boldsymbol{M}$-dependent contribution to the reflected intensity $I = I_0 + I_{ex}$. If $\boldsymbol{P}$ is reversed, V_{ex} and I_{ex} change sign while the polarization-independent contribution I_0 is unaffected. The difference of the intensities $I^{\pm} = I_0 + I_{ex}^{\pm}$ of images taken with $\pm\boldsymbol{P}$, usually normalized with the sum of $I^{\pm}$ and with the degree of polarization $P = |\boldsymbol{P}| \leqslant 1$, $(I_{ex}^{+} - I_{ex}^{-})/(I^{+} + I^{-})P = I_{ex}^{+}/I_0 P = A_{ex}$, is called the exchange asymmetry, and gives an image of the $\boldsymbol{M}$ distribution in the sample. The direction of $\boldsymbol{M}$ can be easily determined by maximizing A_{ex} by rotating $\boldsymbol{P}$ parallel/antiparallel to $\boldsymbol{M}$. Extraction of the magnitude of $\boldsymbol{M}$ requires an A_{ex} analysis in terms of a dynamical SPLEED theory or an empirical calibration.

SPLEEM is particularly useful for the study of crystalline ferromagnetic materials. These have a spin-dependent band structure, with majority-spin and minority-spin bands usually separated by a few tenths of an electron volt (eV) to about 1 eV. Figure 1 shows such an exchange-split band structure [6]. The

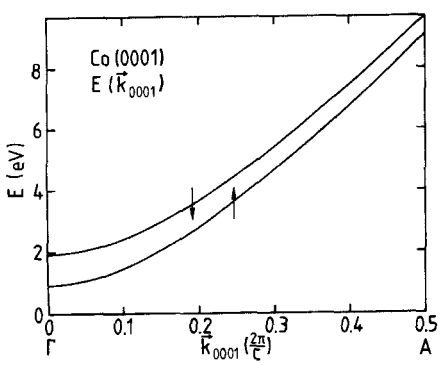


Figure 1. Band structure of cobalt above the vacuum level in the [0001] direction. The Fermi energy is 5.3 eV below the vacuum level [6].

[0001] direction is a frequently encountered orientation in cobalt layers. Below the two energy bands is a large energy gap. Electrons with energies in such a gap cannot propagate in the crystal and are totally reflected (see also Chap. 4, Sec. 1.6 of this Volume). This is true for majority-spin electrons up to 1 eV, and for minority-spin electrons up to 2 eV. Between these two energies there is an increasing excess of minority-spin electrons in the reflected intensity because majority-spin electrons can penetrate into the crystal. Thus, A_{ex} and, therefore, magnetic contrast are large. Above 2 eV, A_{ex} rapidly decreases because both types of electron can now penetrate into the crystal. This is clearly seen in Fig. 2a, in which the intensity I_{00} and measured exchange asymmetry PA_{ex} of the specularly reflected beam from a thick [0001]-oriented single crystalline cobalt layer are plotted as functions of energy [7]. One of the currently most important fields in magnetic materials is the study of ultrathin magnetic layers. These layers frequently show pronounced quantum size effect (QSE) oscillations in A_{ex}. They can be understood by inspecting Fig. 1: to every energy E above 2 eV there are two $\boldsymbol{k} \equiv \boldsymbol{k}_{\perp}$. Although $\boldsymbol{k}_{\perp}$ is not a good quantum number in very thin films, it is still defined well enough to allow thickness-dependent standing waves in the layer, which occur at different thickness for fixed E or at different E for fixed thickness for majority-spin and minority-spin electrons, causing oscillations and a significant enhancement of A_{ex} as seen in Fig. 2b. Such standing waves in the layer are formed not only in energy regions in which the substrate has a band gap but generally whenever there is poor matching of the wave functions at the interface. The

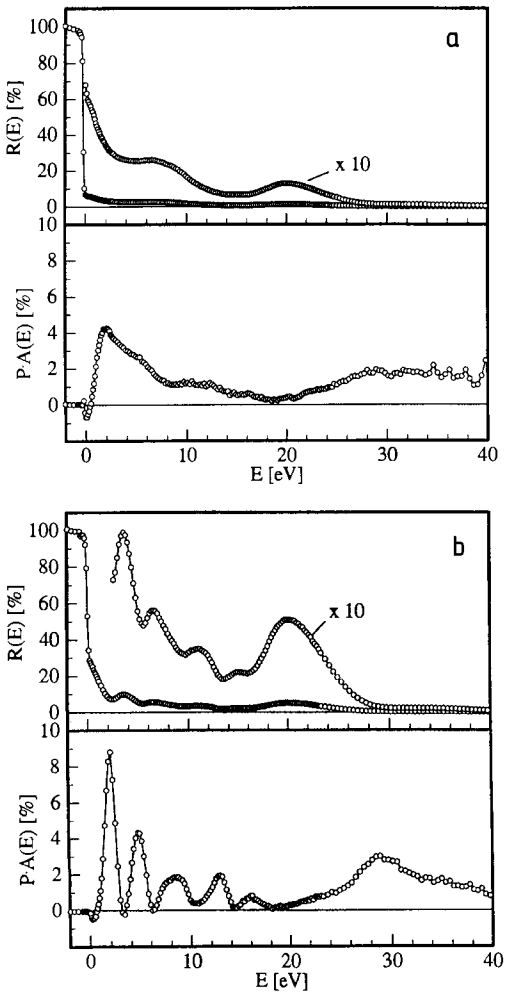


Figure 2. Specular intensity and exchange asymmetry of [0001]-oriented cobalt layers. (a) Thick layer and (b) six monolayers on a W(110) surface [7].

QSE effect in A_{ex} makes determination of the magnitude of $\boldsymbol{M}$ difficult, but is very helpful for contrast enhancement.

The information depth of SPLEEM is the same as in LEEM if spin-dependent inelastic scattering is neglected. Then the inverse attenuation length of electrons with energies of a few electronvolts above the vacuum level in d metals is $1/\lambda \approx 1 + 0.8(5 - n)\,\mathrm{nm}^{-1}$, where n is the number of occupied d orbitals of one spin state [8]. Thus, there is no universal behavior but rather a pronounced material dependence. For cobalt ($n = 4.1$) $\lambda \approx 6$ Å, for iron ($n = 3.5$) $\lambda = 4.5$ Å, while for gold ($n = 5$) $\lambda = 10$ Å. In addition to the limitation by inelastic scattering, the information depth is strongly limited in bandgaps, for example on Co(0001) below 1 eV.

3.3 Instrumentation

A SPLEEM instrument is very similar to a LEEM instrument (see Chap. 4, Sec. 1.6 of this Volume) but differs in the illumination system and in the cathode objective lens. The magnetic field of most magnetic lenses at the specimen position is usually large enough (of the order of 10^{-4} T) to influence magnetic specimens. Therefore, an electrostatic lens is called for, preferentially a tetrode lens which has a resolution comparable to that of a magnetic lens (see Chap. 4, Sec. 1.6 of this Volume). The magnetic sector is unavoidable, but disturbs the polarization of the incident beam only a little because of the low field and short path in it. The illumination system is usually electrostatic, and incorporates the polarization manipulator which is needed to adjust $\boldsymbol{P}$ collinear or normal to $\boldsymbol{M}$. In the present instruments [9, 10] it consists of a crossed $\boldsymbol{B}$–$\boldsymbol{E}$ field 90° deflector and a magnetic field lens as schematically shown in Fig. 3a [11]. The actual configuration of the illumination system is shown in Fig. 3b [10].

The spin-polarized electrons are produced by photoelectron emission from GaAs by left or right circular polarized light from a diode laser. The GaAs(100)

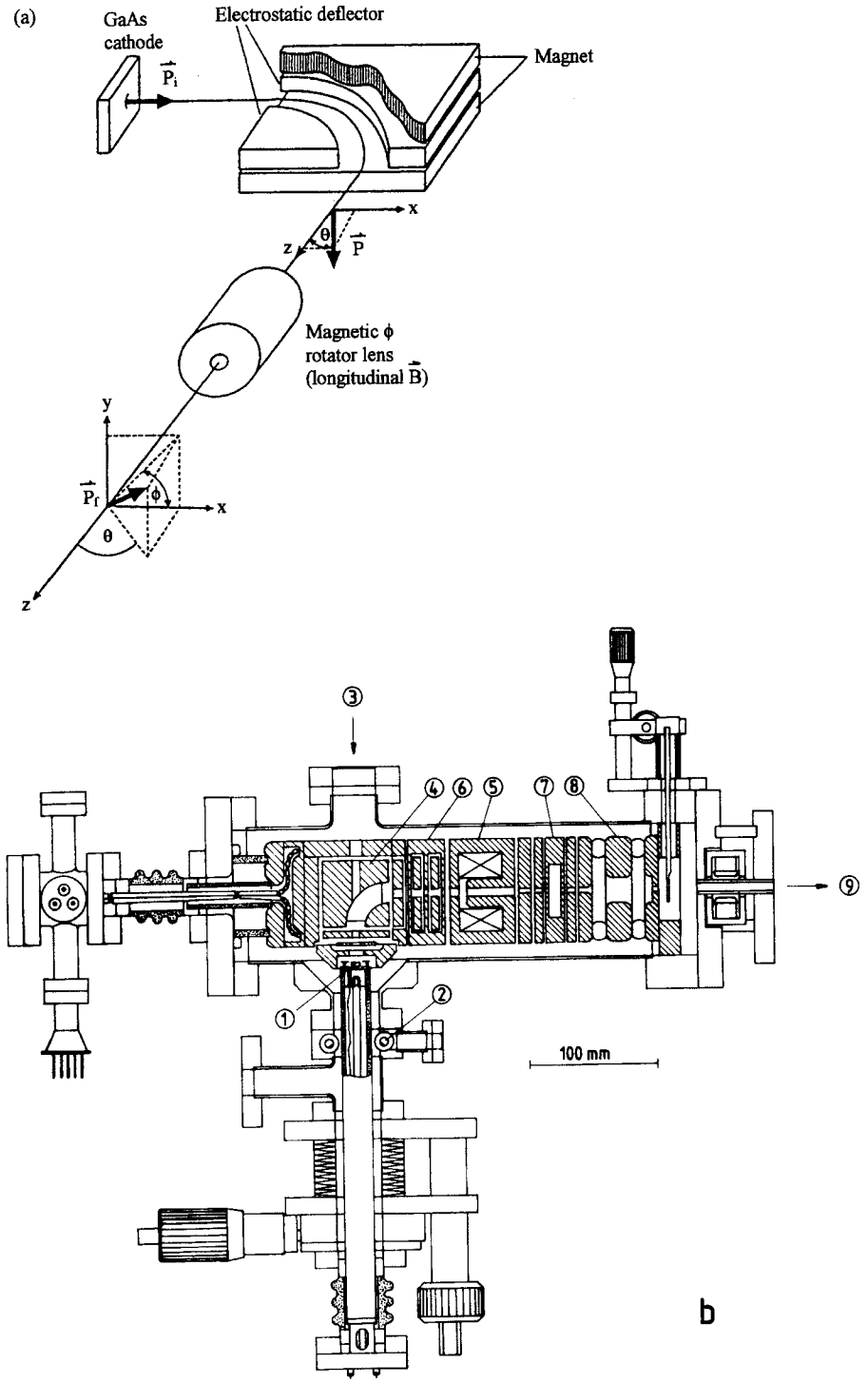


Figure 3. Polarization manipulator. (a) Functional scheme [9]. (b) Configuration with 90° sector and a magnetic field lens at high voltage [10]: 1, GaAs cathode; 2, cesium dispensers; 3, laser beam; 4, ***E–B*** field 90° deflector; 5, magnetic rotator lens; 6, double deflector; 7, double condenser; 8, immersion transfer lens; 9, to beam separator.

single crystal surface is covered with a Cs–O layer in order to approach a negative electron affinity surface. From this surface the electrons are excited into the spin–orbit–split conduction band of GaAs, from which they can escape without having to overcome a barrier. The maximum theoretical polarization obtainable from such an emitter is 50% but in practice 20–25% is usual. Contrary to the LaB_6 emitter used in LEEM which has thousands of operating hours, the activation of GaAs emitters has a useful life of the order of hours to days depending upon the vacuum, and activation is still more an art than a science. Strained alloy multilayers on GaAs basis are now available with much higher polarizations, but the current and reliability are still inferior to the standard GaAs emitter.

The resolution in SPLEEM is comparable to that in LEEM, but the signal-to-noise ratio is poorer due to the need for image subtraction if purely magnetic images are desired. Inasmuch as $I_{ex} \ll I_0$, longer exposures (in the 1 s range) are needed for a good signal-to-noise ratio. Further instrument development should allow image acquisition times of the order of 0.1 s. The field of view is up to about 50 μm, similar to LEEM.

3.4 Areas of Application

The magnetic domain structure on the surface of bulk magnetic materials is generally too large grained for SPLEEM studies. Exceptions are, for example, closure domains on Co(0001) surfaces or domain walls [12]. The major field of SPLEEM is, therefore, the study of thin films and superlattices whose magnetic properties depend strongly upon film thickness and structure, in particular upon interface structure.

One of the questions is to what extent the substrate influences the magnetic properties. Cobalt layers on W(110) illustrate this influence. On clean W(110), cobalt grows with the closest-packed plane parallel to the substrate, with h.c.p. stacking, at least at seven monolayers ((0001) orientation), on the W_2C-covered surface in f.c.c. packing with the (100) orientation [13]. In both cases the magnetization is in-plane and has a pronounced uniaxial anisotropy with the easy axis in the h.c.p. layer parallel to W[110] [14]. In both cases good magnetic contrast is obtained at three monolayers, although the magnetic moment is still small at this thickness. The magnetic domains are initially small but rapidly coalesce with increasing film thickness [13]. Steps on the substrate surface have no influence on the size and shape of the domains and apparently neither on the location of the vertices in the Néel-type domain walls which can be imaged with $\boldsymbol{P}$ perpendicular to $\boldsymbol{M}$. Sometimes the vertices are located at steps seen in the unsubtracted images of Fig. 4, but more often they occur on terraces [15]. In contrast to cobalt on W(110), which is in-plane magnetized from the very beginning, cobalt on Au(111) has initially out-of-plane magnetization which switches to in-plane at 4–5 monolayers. The exact thickness at which this transition occurs depends upon the state of the Au(111) surface. In particular, when cobalt is grown on an Au(111) double layer on W(110), the step distribution can have a strong influence on domain size and shape [16a].

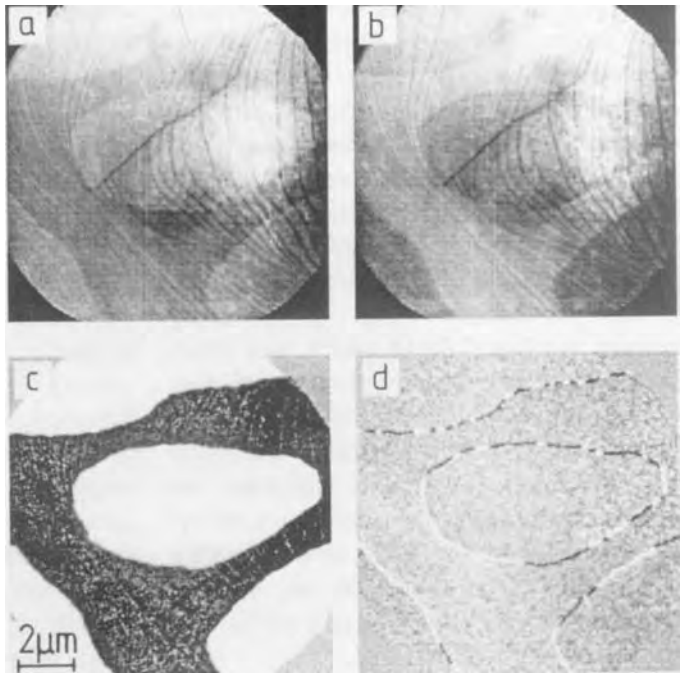


Figure 4. Magnetic domain images of a six monolayer thick cobalt layer on W(110). Images with ***P*** (a) parallel and (b) antiparallel to ***M***, respectively. (c) Difference image between (a) and (b). (d) Difference image between two images with ***P*** perpendicular to ***M***. Electron energy 2 eV [15].

The in situ study capabilities of SPLEEM instruments (vapor deposition, heating, cooling, etc.) also allow the study of changes in the magnetic structure as a function of film thickness or temperature. For example, during growth of cobalt on clean W(110) at an elevated temperature, for example 700 K, flat three-dimensional single-domain cobalt crystals form on a cobalt monolayer. When the crystals grow to join each other with increasing thickness, the domains rearrange, which can be followed by SPLEEM quasi-life [14]. Another example is the change in magnetization when a continuous cobalt layer breaks up into small crystals during annealing [14].

Nonmagnetic overlayers on magnetic layers can have a strong influence on their magnetic structure. An example is gold on Co(0001) layers. Figure 5a is the SPLEEM image of a 10 monolayer thick

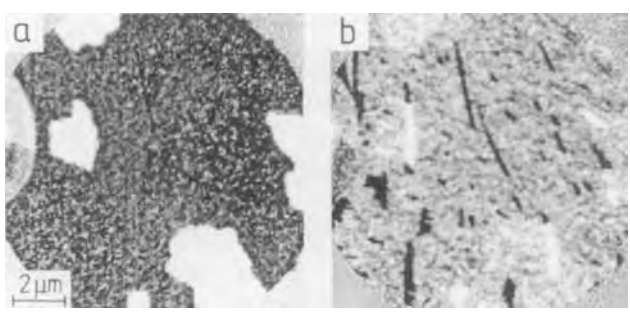


Figure 5. Influence of a gold overlayer on the magnetic structure of a 10 monolayer thick cobalt layer on W(110). (a) SPLEEM image of the cobalt layer and (b) SPLEEM image of the same area covered with two monolayers of gold. Electron energy 1.5 eV [16b].

cobalt layer deposited on W(110) at 400 K. One monolayer of gold deposited onto this cobalt layer at room temperature has no influence on the domain structure, but when two gold monolayers are deposited, most of the cobalt layer switches to out-of-plane magnetization—so that $\boldsymbol{M} \perp \boldsymbol{P}$ and $A_{ex} = 0$—but on some terraces the original in-plane magnetization remains (Fig. 5b), even if additional gold is deposited [16b]. Figure 5 illustrates very well how the topography of the substrate can be propagated through a deposited layer and influence its magnetic properties. In contrast to gold, copper overlayers have no influence on the domain structure in cobalt layers up to the largest overlayer thickness studied (14 layers). Only pronounced QSE A_{ex} oscillations are seen as a function of copper layer thickness [17].

One of the most interesting subjects in magnetism has been for some time the magnetic coupling between magnetic layers through nonmagnetic layers. Frequently it oscillates between ferromagnetic (FM) and antiferromagnetic (AFM) coupling, and the causes of the AFM coupling and its occasional absence has been the subject of much discussion. Here, SPLEEM can make an important contribution because of its significant information depth at very low energies. As an example, Fig. 6 shows SPLEEM images of a cobalt layer on W(110) and of a cobalt–copper–cobalt sandwich grown on it by sequential deposition of four monolayers of copper and six monolayers of cobalt. According to most magnetic studies, the two cobalt layers should be antiferromagnetically coupled through four monolayers of copper. Analysis of many SPLEEM images similar to Fig. 6, however, shows that there is only locally occasional AFM coupling, but also FM

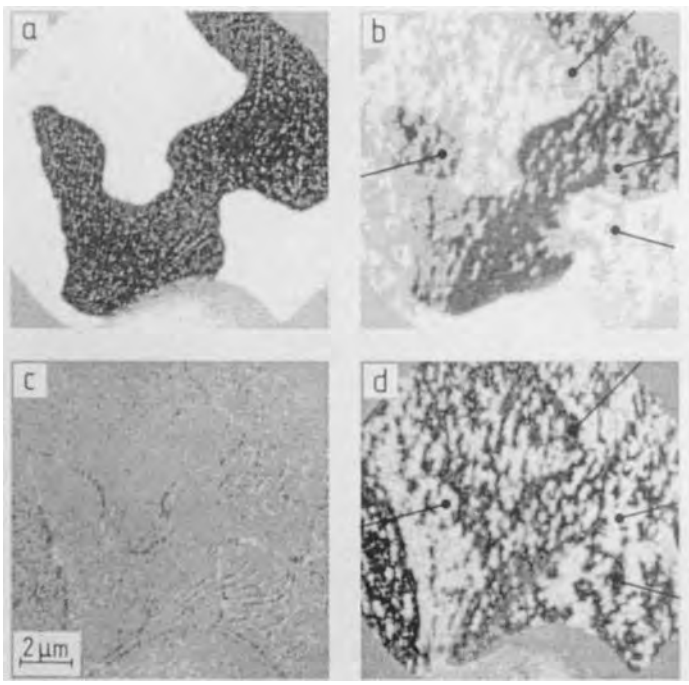


Figure 6. Magnetic coupling between two cobalt layers through a copper layer. (a, c) Bottom cobalt layer only and (b, d) complete cobalt–copper–cobalt sandwich. $\boldsymbol{P} \| \boldsymbol{M}$ in (a) and $\boldsymbol{P} \perp \boldsymbol{M}$ in (b) in bottom layer. The solid circles denote identical domains in the top layer. Electron energy 1.5 eV [15].

coupling, and other $\boldsymbol{M}$ directions in the top layer, which also has much smaller domains [15]. These domains extend frequently across the boundaries of the domains in the lower cobalt layer which can be seen through the copper and the top cobalt layer.

3.5 Discussion

When should SPLEEM be used for the study of the magnetic microstructure of materials, and when not? Section 3.4 of this Chapter has already mentioned a number of interesting applications, but there are many more such as the study of magnetic phase transitions as a function of temperature, magnetic switching phenomena in pulsed fields, or magnetization processes in fields perpendicular to the surface. Fields parallel to the surface can only be applied in a pulsed manner because of the beam deflection which they cause, while fields normal to the surface require only refocusing. The areas for which SPLEEM is not well suited are essentially the same as in LEEM (see Chap. 4, Sec. 1.6 of this Volume): rough surfaces, high vapor pressure materials, etc.

There are many other magnetic imaging methods. Most of them do not image magnetization but the internal or external magnetic field distribution caused by the magnetization distribution such as Lorentz microscopy (see Chap. 4, Sec. 1.8 of this Volume), electron holography (see Chap. 4, Sec. 1.9 of this Volume), or magnetic force microscopy (see Chap. 7, Sec. 3 of this Volume). These techniques are to a large extent complementary to SPLEEM, as is MEM [18], which can be easily combined with LEEM. The most important competitor to SPLEEM is scanning electron microscopy with polarization analysis (SEMPA; see Chap. 5, Sec. 2 of this Volume), which images the magnetization distribution in the specimen via polarization analysis of the secondary electrons. The advantages of SEMPA over SPLEEM are an easy combination with electron spectroscopy, unimportance of the crystallinity of the specimen for high brightness, and the absence of a high electric field at the specimen surface. Some of the advantages of SPLEEM over SEMPA are rapid image acquisition, and easy combination with LEED and various emission microscopies.

3.6 Concluding Remarks

At present, SPLEEM is still too young a technique to allow an extensive discussion of its advantages and disadvantages. Only two instruments are in operation at the time of writing. Commercial instruments will probably not be available before 1997.

3.7 References

[1] J. Kessler, *Polarized Electrons*, 2nd ed., Springer, Berlin **1985**.
[2] D. T. Pierce, R. J. Celotta, *Adv. Electron. Electron Phys.* **1981**, *65*, 219.
[3] R. Feder, *J. Phys.* C, **1981**, *14*, 2049; R. Feder, *Phys. Ser.* **1983**, *74*, 47.
[4] R. Feder (Ed.), *Polarized Electrons in Surface Physics*, World Scientific, Singapore **1985**.
[5] J. Kirschner, *Polarized Electrons at Surfaces*, Springer, Berlin **1985**.
[6] J. Noffke, personal communication.

[7] K. Wurm, M.S. thesis, TU Clausthal **1994**.
[8] H. C. Siegmann, *J. Phys.: Condens. Matter* **1992,** *4*, 8395.
[9] K. Grzelakowski, T. Duden, E. Bauer, H. Poppa, S. Chiang, *IEEE Trans. Mag.* **1994,** *30*, 4500.
[10] T. Duden, Ph.D. thesis, TU Clausthal 1996.
[11] T. Duden, E. Bauer, *Rev. Sci. Instrum.* **1995,** *66*, 2861.
[12] M. S. Altman, H. Pinkvos, J. Hurst, H. Poppa, G. Marx, E. Bauer, *MRS Symp. Proc.* **1991,** *232*, 125.
[13] H. Pinkvos, H. Poppa, E. Bauer, J. Hurst, *Ultramicroscopy* **1992,** *47*, 339.
[14] H. Pinkvos, H. Poppa, E. Bauer, G.-M. Kim in *Magnetism and Structure in Systems of Reduced Dimensions* (Eds.: R. F. C. Farrow, B. Dieny, M. Donath, A. Fert, B. D. Hermsmeier), Plenum, New York **1993,** p. 25.
[15] E. Bauer, T. Duden, H. Pinkvos, H. Poppa, K. Wurm, *J. Magn. Magn. Mat.*, in press.
[16] (a) M. S. Altman, H. Pinkvos, E. Bauer, *J. Magn. Soc. Jpn.* **1995,** *19* (Suppl. S1), 129; (b) M. S. Altman, H. Pinkvos, E. Bauer, unpublished findings.
[17] H. Poppa, H. Pinkvos, K. Wurm, E. Bauer, *MRS Symp. Proc.* **1993,** *313*, 219.
[18] H. Bethge and J. Heydenreich, *Elektronenmikroskopie in der Festkörperphysik*, Springer, Berlin **1982,** p. 196.

Part VI

Emission Methods

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

1 Photoelectron Emission Microscopy

1.1 Introduction

Photoelectron emission is the ejection of electrons from matter by the action of light. Once released from matter, photoelectrons become indistinguishable from electrons produced by other means. In particular, photoelectrons emitted into vacuum are deflected by electric and magnetic fields and a beam of photoelectrons can be focused by electrostatic and electromagnetic lenses to form a photoelectron image of a sample. This is the basis of photoelectron microscopy.

1.2 Photoelectron Emission

The origins of research into photoelectron emission are traditionally traced to a report by Heinrich Hertz that ultraviolet light increased the length of an electric spark in the spark gap of an electric oscillator [1, 2]. Hertz was in the process of discovering radio waves, which he detected by observing sparks in a secondary electric oscillator which was in resonance with a primary oscillator powered by an induction coil. He noticed that the sparks of the secondary were more intense if the spark gap had direct line-of-sight to the spark gap of the primary. Using ultraviolet filters, he showed that ultraviolet light produced by the primary spark was responsible for the effect. Furthermore, he showed that the effect was produced by using other sources of ultraviolet light such as a carbon arc lamp.

The next advance was initiated by Hallwachs, who used a gold-leaf electroscope in air to show that negative charge was emitted from zinc exposed to ultraviolet light [3–5]. An insulated, uncharged sample acquired a positive charge, and a negatively charged sample lost its negative charge and became positively charged. However, a strongly positively charged sample retained its charge. Hallwachs suggested that the ultraviolet light caused the emission of negatively electrified particles [5]. A high positive potential on the sample prevented emission of negative particles by electrostatic attraction.

Elster and Geitel showed that photoelectric rays in a cathode ray tube were deflected by a magnet and behaved like other cathode rays [6, 7]. Using crossed electric and magnetic fields, Thomson [8, 9] and Lenard [9, 10] showed that the charge-to-mass ratio [11] of the negative charge carriers of photoemission was identical to that of electrons. Lenard showed that the photoemitted current (i.e., the number of electrons released)

increased with the intensity of light. However, the kinetic energy of the photoelectrons was independent of the light intensity. Moreover, the kinetic energy of the photoelectrons increased with the increase in the frequency of light. Furthermore, photoelectrons were not ejected from a sample unless the frequency of the incident light was greater than a threshold value which was characteristic for each substance.

Einstein [12, 13] explained Lenard's results by making the following assumption: matter contains bound electron resonators which emit and absorb electromagnetic waves of a definite period. The resonators emit light with a specific quantum of energy, $E = h\nu$ (in modern terms) where h is Planck's constant and ν is frequency of light. Unlike water waves which have a decreasing wave amplitude and eventually become imperceptible, electromagnetic waves do not dissipate as they propagate. Rather, the energy emitted, $h\nu$, remains intact in a light quantum, which can deliver its entire energy to a single electron. This energy can be transformed into the kinetic energy of an electron. Further, if it is assumed that the electron is bound to matter by some potential energy, P, the electron cannot be emitted unless the quantum of light possesses at least this binding energy. Perhaps the most convincing evidence evoked by Einstein to show that light does not spread as does a water wave is the fact that far ultraviolet light photoionizes single atoms and molecules in the gas phase [14]. Because gas molecules have cross sections of only a fraction of a few square nanometers, the quantum of light cannot extend over an area much greater than atomic dimensions.

It does not follow from Einstein's theory of photoemission that light is a billiard ball like corpuscle. Light must be a form of motion for the same reason that heat is considered to be a form of motion. In his classic studies of heat, Rumford showed that water could be boiled and heat could be produced indefinitely by friction. Because the heat produced was inexhaustible, Rumford concluded that heat could not be a material substance but must be a form of motion [15]. In taking Rumford's experiment a step further, a sample can be warmed by friction until it glows with a red heat. Like heat, an inexhaustible supply of photons is produced from matter by a hot glowing body. Photons, therefore, cannot be a material substance released from matter, but must be a form of motion. This motion can be transformed into the kinetic energy of a photoelectron.

Our most detailed knowledge of the events involved in photoelectron emission is revealed by photoelectron spectroscopy [16–19]. In this technique, the kinetic energy of photoelectrons can be analyzed at meV resolution. The following fact has emerged: photoelectrons can be ejected from all atoms, molecules, liquids, and solids if photons of sufficient energy to overcome the electron binding energy are absorbed. Electrons within isolated atoms and molecules in the gas phase are bound in atomic and molecular orbitals, each with a characteristic binding energy [18]. For most gases, including organic molecules, first ionization potentials are rarely below 5 eV and more generally start between 9 and 13 eV and require vacuum ultraviolet radiation for photoionization [18]. More strongly bound core electrons need X-ray energies for emission [17]. In

addition to overcoming the potential energy which binds the photoelectron, photoemission may be accompanied by the excitation of the photoionized molecule to higher translational, vibrational, rotational and electronic states [18]. Thus the kinetic energy of the photoelectron can be reduced by these loss mechanisms, and a spread of kinetic energies is expected even if using a monochromatic light source. Conversely, if the molecule initially is in an excited state, the translational, vibrational, rotational, or electronic energy can be added to the kinetic energy of the photoelectron, albeit usually with a lower probability than in the case of energy loss.

Solid metals and semiconductors generally have threshold energies below 6.5 eV, which is much lower than typical isolated molecules. For these materials, near ultraviolet light is sufficient to induce photoemission. Films of strontium, barium, sodium, potassium, rubidium, cesium and lithium, distilled onto silver coatings on glass will emit with visible light, and were used in early photocells [20]. The remarkable material Ag–O–Cs, has a photoelectric threshold near 1400 nm [7] (p. 93). It was commercialized in the 1930s, but remains one of the very few materials which photoemits in the near infrared [21]. A roughened surface of silver is essential for emission [21].

Once ejected from an atom within a solid, the photoelectron becomes subject to the same forces which dictate motion of all electrons in matter. Photoelectrons undergo electron diffraction [22] and can propagate only with energies and in crystallographic directions allowed by band theory. Photoelectrons are subject to the same energy loss mechanisms as other electrons and can lose kinetic energy to lattice vibrations (phonons), plasmons, or electronic excitations. Electrons in semiconductors can be excited across the band gap, and very energetic photoelectrons can induce a secondary electron cascade. The mean-free-path of a photoelectron in matter depends upon these loss mechanisms. For photoelectrons of energies from 1–2 eV, the mean-free-path can be large (100 nm) [19]. This is because electrons at very low energies are not energetic enough to excite loss mechanisms, which require a definite minimum quantum of energy. Electrons of 100 eV energy have much shorter paths of only a few nanometers [19].

To be emitted, a photoelectron originating in the bulk must reach the surface with sufficient energy to overcome the surface potential barrier, which in metals is the photoelectric work function. The work function varies between materials and crystallographic planes of a material. The work function can be drastically changed by absorption of a single monolayer [19]. Photoelectric threshold and photoelectric intensity depend critically on the state of the surface and the nature of the adsorbate. For this reason, it is possible to image single adsorbed monolayers, including submonolayer coverage, by photoelectron microscopy.

1.3 Microscopy with Photoelectrons

Photoelectron emission microscopy dates to the very beginnings of electron microscopy. Brüche [23, 24] published a photoelectron image of zinc in 1933, just one

year after publication of the first thermionic electron emission images by Brüche and Johannson [25] and transmission electron images by Knoll and Ruska [26]. Other early papers showed that the technique was sensitive to the adsorption of gases on platinuum [27, 28] and nickel [29]. It was proved that metals such as potassium and barium could be deposited in situ onto metallic substrates, and that their adsorption (presumably of a single monolayer) caused a great increase in photoelectron emission by reduction of the work function. Several reviews are written on photoelectron microscopy [30–37]. In traditional photoelectron emission microscopy, flat solid samples are flooded with light of sufficient energy to overcome the work function threshold of the sample. Commonly used is the low pressure mercury lamp (Osram HBO 100W/2). Spectra of Hg [38] show sharp peaks of various intensity at 234.54 nm (5.29 eV, very faint), 239.97 nm (5.17 eV, faint), 244.69 nm (5.07 eV, faint), 246.41 nm (5.03 eV, faint), 248.27 nm (5.00 eV, moderate intensity) and 253.65 nm (4.89 eV, the strongest). Many other peaks exist at lower energies [38–41]. Quartz optics and windows transparent to UV light are necessary. The user needs to be aware that samples with work functions above 5.3 eV will not be imaged using a mercury lamp. Oxide and carbon monoxide contamination often raises the work function of samples to above 5.3 eV. Conversely, ubiquitous carbon contamination can lower the work function to below 4 eV.

Deuterium lamps, which have a maximum energy of 6.9 eV, are used to eject photoelectrons from samples with higher work function such as platinum or oxygen and carbon monoxide adsorbed on platinum. Sapphire windows transparent to UV are used, and oxygen must be purged from the beam line to avoid production of ozone and adsorption of UV light by oxygen. Synchrotron radiation [42] has been used successfully in photoelectron imaging [36, 43–46]. Attempts have been made to use conventional X-ray sources [47], and lasers [48], but these are not commonly used. Xenon lamps have also worked. In conventional nonscanning photoelectron microscopy, light need not be focused to a small spot size. However, some focusing is desirable to increase intensity. For Hg lamps, spots sizes are typically greater than 3 mm in diameter.

Photoelectrons in vacuum are accelerated and focused with classical electron optics, such as the electrostatic triode and tetrode objective lenses and the magnetic objective lens drawn in Fig. 1. Most instruments employ additional intermediate and projector lenses similar in principle to those used in transmission electron microscopes. The electron optics of such lenses is discussed elsewhere [30, 33, 34, 49–60]. Channel plate image intensifiers and ultrahigh vacuum conditions are used in most advanced instruments.

The early triode lens of Brüche and Johannson was made simply of two flat parallel metal plates spaced 1.0 mm apart with a 1.2 mm hole bored into the plate nearest the sample and a 1.0 mm hole in the second [55]. A flat sample, negative with respect to the plates, formed the third electrode of the triode. Focusing was achieved by varying the potential of the plate nearest the sample. In the advanced version (that of E. Bauer [34]) of the triode lens of Fig. 1, the metal plates are sloped to allow light to reach the sample ($\approx 75°$ off normal) and to allow

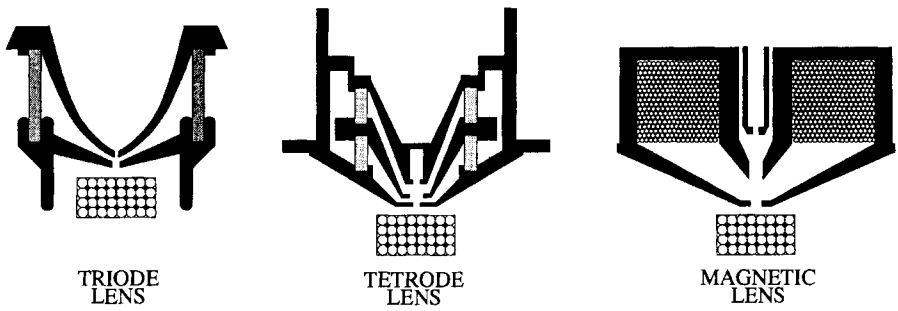


Figure 1. The three main types of objective lenses used in photoelectron emission microscopy. In the electrostatic triode and tetrode lenses, metallic regions (dark) are separated by ceramic insulators (grey). Flat photoemitting solids are placed ~4 mm beneath the lenses and are illuminated with light of sufficient energy to overcome the photoelectric threshold of the sample. Samples can be thick (many centimeters) and can be heated or cooled from beneath.

insertion of a wider ceramic insulator (shaded) to avoid voltage breakdown. The triode lens has a proven resolution of 15 nm [61]. The tetrode electrostatic objective lens developed by Johannson in the early 1930s [55], consisted of three flat parallel plates. The flat sample formed the fourth electrode of the tetrode. In the advanced version (that of W. Engel [88]) of Fig. 1, the plates are again sloped away from the optical axis to allow the beam of light to reach the sample and to allow insertion of wider insulators. Focusing is achieved by varying the potential of the central electrode. Proven resolution is better than 10 nm [62]. A disadvantage of the tetrode is that the entire accelerating field (up to 30 keV over 1–4 mm) is between the sample and nearest electrode. Voltage breakdown between sample and first electrode can occur if the sample is not smooth. In the triode lens, the voltage between sample and focusing electrode is only a few hundred volts over a few millimeters. Resolution is lost, but voltage breakdown is less likely. The early photoelectron microscopes used simple coils of copper as electromagnetic lenses [23, 27]. The version of the magnetic lens of Fig. 1, developed by W. Engel, has a resolution of better than 10 nm [34]. It is more robust than the electrostatic lenses, which must best be used in ultrahigh vacuum to avoid deposition of conducting carbonaceous material onto the insulators. The magnetic field, however, can interfere with the study of magnetic materials. The coils of the lens need to be cooled to avoid heating and degassing of the lens. Resolution is limited in practice to chromatic aberrations due to the kinetic energy spread of the emitted photoelectrons [50]. Optimum resolution is achieved only if the energy spread is less than 0.5 eV. Thus best resolution is obtained if the light energy is only a few tenths of an electron volt above threshold energy. For a given instrument, resolution will be highly sample dependent. Although it may seem trivial, mechanical vibrations from pedestrian and motor traffic can easily degrade resolution into the micron range. Stray electromagnetic fields, for example, from fluorescent lights in the same building (even a few floors away) or from ionization gauges, etc., can also degrade resolution into the micron range.

1.4 Applications

1.4.1 Monolayer Epitaxial Growth

Figure 2a shows the initial stages of the epitaxial growth of the first monolayer of copper evaporated onto an Mo {0 1 1} single crystal surface. To avoid contamination by residual gas, the system had a base pressure of 2×10^{-10} torr, and the film was grown in situ at a rate of one full monolayer in 2.5 min [61]. The substrate was at 700 K which allowed the copper to diffuse to atomic steps which were present. Figure 2b shows continued growth of the first monolayer in another region having wider terrace widths. The curved step

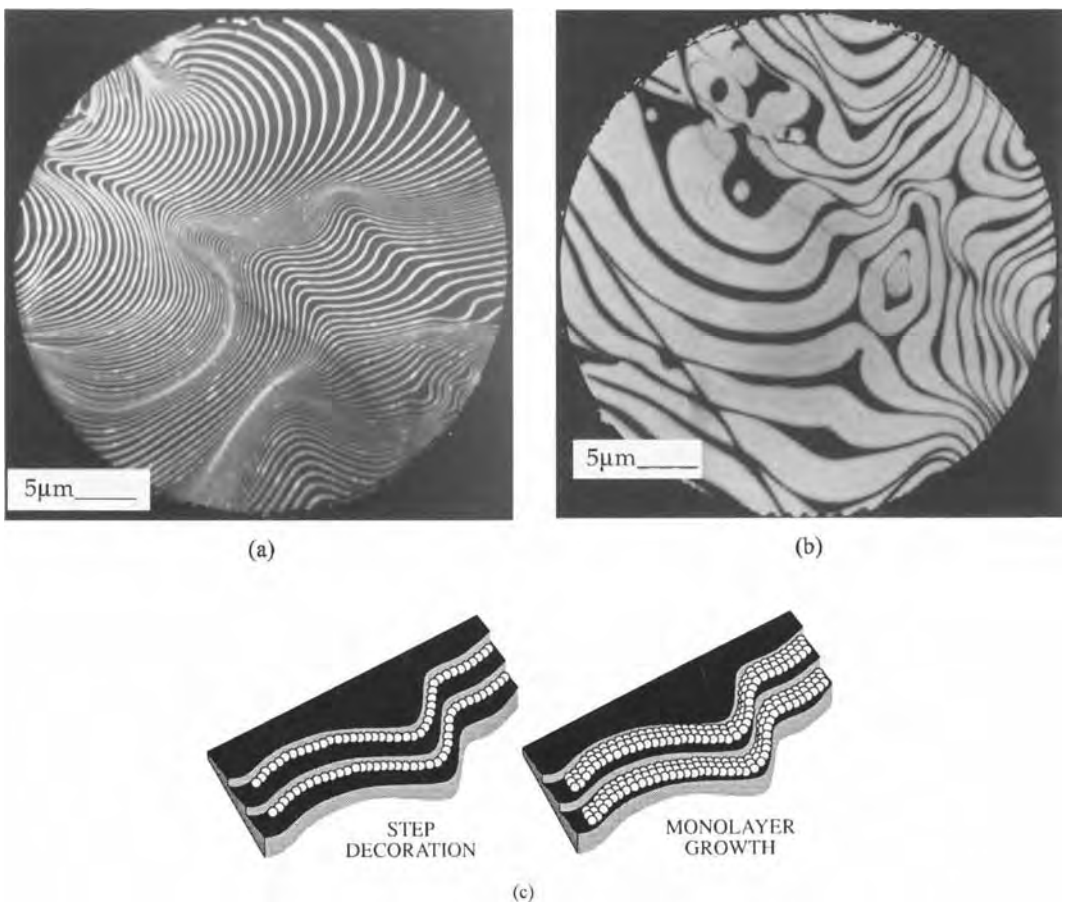


Figure 2. (a) Initial growth of copper monolayer at atomic steps of an Mo {0 1 1} surface at 700 K. Curved steps of Mo arise by step migration during substrate preparation. (b) Continued first monolayer growth of Cu/Mo {0 1 1} under step flow growth conditions. Near perfect contrast in monolayer growth is achieved by choosing a substrate with threshold greater than the light source and an adsorbate with a lower work function. (c) Schematic model of adsorption and growth at atomic steps. In Fig. 2a the white bands are one monolayer thick and ~250 copper atoms wide.

structure was due to step migration during preparation and cleaning of the molybdenum. Details are explained elsewhere [61, 63–69]. Atomic steps of copper form at the steps of molybdenum, as shown schematically in Fig. 2c, and monolayer growth follows the classical step-flow model of crystal growth [70]. An Hg lamp and triode lens were used. The clean Mo {0 1 1} face has a work function above 5.3 eV. Its photoelectric threshold is, therefore, above the maximum transmitted energy of the Hg lamp. No photoelectrons are emitted from the uncovered molybdenum, which appears absolutely black. Adsorbed copper lowers the work function to 4.8 eV [61]. Photoelectrons are emitted from the copper covered areas and appear bright. By choosing a substrate with work function above the maximum energy of the light source, and by choosing an adsorbate with a threshold within the range of the lamp, single monolayers are imaged and with perfect contrast.

The photoelectrons emitted from the copper covered areas can originate from the underlying Mo as well as from the copper atoms. By reducing the surface potential, photoelectrons from Mo can escape if they have kinetic energy, greater than 4.8 eV. Because the energy spread does not exceed 0.5 eV, optimal resolution of 15 nm is achieved with a triode lens. Use of photoelectron microscopy to study epitaxial growth is described elsewhere [61, 63–73]. It should be pointed out that the crystal face of Fig. 2 was especially well oriented (0.03° off axis). This gave exceptionally wide terraces. Samples with terrace widths smaller than the resolution limits will not give such results.

1.4.2 Chemical Kinetic Reaction-Diffusion Fronts in Monolayers

Study of spatio-temporal pattern formation and temporal oscillations in chemical systems is now at the forefront of chemical kinetics and chemical physics [74–78]. Chemical waves, target patterns, spirals, and complex 'chemical chaos' are observed in reactive systems and can be mathematically modeled by solving nonlinear differential equations [74–78]. Such phenomena occur in reactive monolayer films on catalyst surfaces and can be observed directly by photoelectron microscopy [79–87]. In fact, no other microscopic technique has revealed these phenomena in monolayers in such detail as has photoelectron microscopy. Figure 3 shows spiral formation seen during the catalytic reaction between carbon monoxide and oxygen on Pt {1 1 0} at 435 K. The darker areas are covered with an adsorbed monolayer of oxygen and the gray areas are covered with a monolayer of adsorbed carbon monoxide. Such spirals are also seen in liquid systems and arise from the chemical kinetics and *not*, for example, from screw dislocations on a surface. A still photograph cannot do justice to the myriad of dynamic events recorded at video rates. Slight changes in reaction conditions can completely alter the spatio-temporal patterns. The image was taken using a deuterium lamp and electrostatic tetrode lens [88].

1.4.3 Magnetic Materials

Ferromagnetic domains have been imaged by photoelectron microscopy [89–92]. The

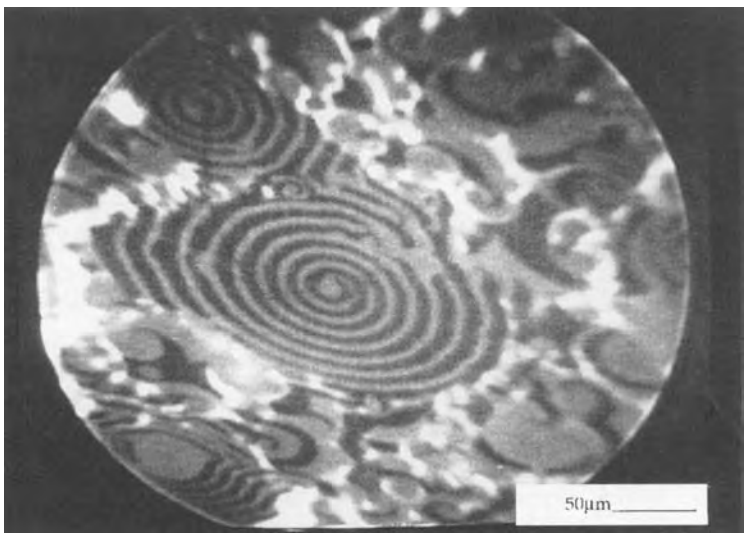


Figure 3. Distribution image of the reactants CO (grey) and O (dark) on Pt {1 1 0} during catalytic reaction. Sample temperature of 435 K and partial pressures of 3×10^{-4} mbar O_2 and 3.2×10^{-5} mbar CO in the reaction chamber. (Courtesy of W. Engel.)

magnetic field of oppositely magnetized domains can split the photoelectron beam into two. As in Lorentz microscopy, one beam can be stopped by an aperture, allowing domains of one orientation to appear bright and those of opposite orientation to be dark.

Using circularly polarized X-rays from a synchrotron source to excite electrons of a specific spin orientation, Tonner et al.

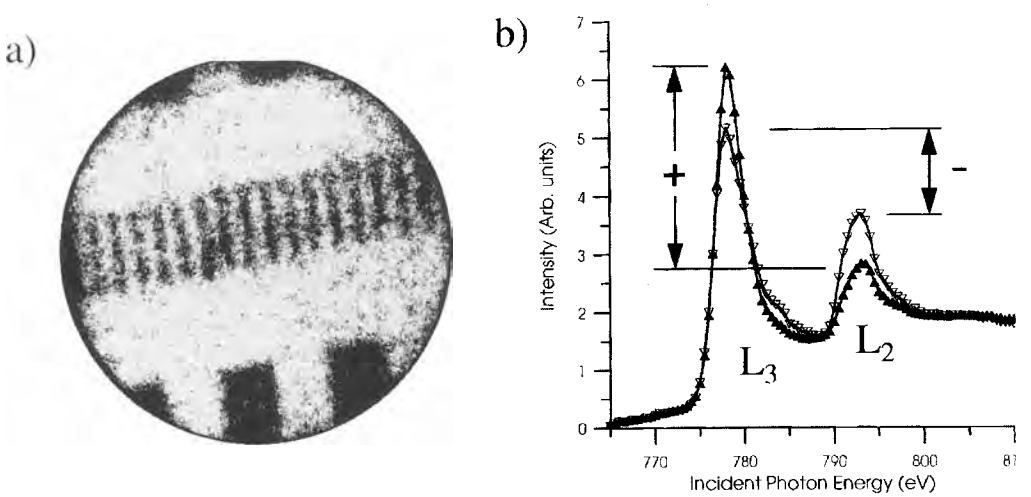


Figure 4. (a) High resolution MCXD image of magnetic domains on a Co–Pt magnetic recording disk obtained using circularly polarized X-rays. The image was recorded by subtracting images obtained at the cobalt L_3 and L_2 edges for enhanced contrast. Image field-of-view: 70 μm. (b) X-ray absorption spectra near the Cobalt L edges of the same sample recorded with circularly polarized X-rays and parallel (filled triangles pointing up) and antiparallel (open triangles pointing down) alignment of photon spin and magnetization vectors, respectively. The contrast in the image is indicated by the difference in height of the + and − arrows. (Courtesy of D. Dunham and B. P. Tonner.)

[93, 94] have imaged magnetized regions of a magnetic recording disc as shown in Fig. 4a. The image was correlated with photoelectron spectra (Fig. 4b). The use of electrostatic objective lenses is one great advantage of photoelectron microscopy in the study of magnetic materials because the field of a magnetic objective lens does not interfere with domain structure of the magnetic material.

1.5 Choice and Preparation of Samples

Because the sample is actually part of the objective lens, samples should be flat to avoid field distortion. Single crystal surfaces, polished metallurgical specimens, silicon wafers and integrated circuits work well. Samples need to conduct well enough to replace the electrons emitted in order to avoid charging. Metals and semiconductors work well. Samples which are normally classified as insulators have been imaged if they become photoconducting or can be heated until they conduct [95]. Insulators can, of course, be coated with a conducting layer if surface sensitivity is not necessary. For controlled experiments and monolayer studies, ultra high vacuum compatibility is essential. Sample surfaces need to be cleaned in situ by heating, chemical reaction or sputtering. Although photoelectron emission microscopy can be used for routine postmortem examination of samples, best results are obtained by performing experiments in situ with the sample in the observational position. These include epitaxial growth and desorption of monolayers, chemical vapor deposition [96], reaction–diffusion phenomena, etc. Samples can be heated and cooled. Thick samples (1–3 cm) and even entire (10 cm) silicon wafers can be used.

1.6 References

[1] H. Hertz, *Ann. Phys.* **1887,** *31*, 983.
[2] H. Hertz, *Electric Waves*, Dover, New York **1962,** p. 63.
[3] W. Hallwachs, *Ann. Phys.* **1888,** *33*, 301.
[4] S. P. Thompson, *Light Visible and Invisible*, MacMillan, London **1928,** p. 181.
[5] T. H. Osgood, A. E. Ruark, E. Hutchisson, *Atoms, Radiation and Nuclei*, Wiley & Sons, New York **1964,** p. 78.
[6] J. Elster, H. Geitel, *Ann. Phys.* **1890,** *41*, 166.
[7] V. K. Zworykin, E. D. Wilson, *Photocells and Their Application*, Wiley, New York, **1932,** p. 9.
[8] J. J. Thomson, *Phil. Mag.* **1899,** *48*, 547.
[9] A. L. Hughes, L. A. DuBridge, *Photoelectric Phenomena*, McGraw-Hill, New York **1932,** p. 4.
[10] P. Lenard, *Ann. Phys.* **1902,** *8*, 169.
[11] O. Klemperer, *Electron Physics, The Physics of the Free Electron*, Academic Press, New York **1959,** p. 121.
[12] A. Einstein, *Ann. Phys.* **1905,** *17*, 132.
[13] A. B. Arons, M. B. Peppard, *Am. J. Phys.* **1965,** *33*, 367.
[14] A. von Engel, *Ionized Gases*, Oxford University Press, Oxford **1955,** p. 60.
[15] B. Thompson, Count Rumford in *A Source Book in Physics* (Ed.: W. F. Magie), Harvard University Press, Cambridge, MA **1969,** p. 161.
[16] K. Siegbahn, C. Nordling, A. Fahlman, R. Nordberg, K. Hamrin, J. Hedman, G. Johanson, T. Bergmark, S. Karlsson, I. Lindgren, B. Lindberg, *Nova Acta Regie Soc. Sci. Upsal.* **1967,** *20*, 4.
[17] K. Siegbahn in *Electron Spectroscopy* (Ed.: D. A. Shirley), North-Holland, Amsterdam **1972,** p. 15.
[18] C. R. Brundle, M. B. Robin in *Determination of Organic Structures by Physical Methods* (Eds.: F. C. Nachod, J. J. Zuckerman), Academic Press, New York **1971,** p. 1.
[19] D. P. Woodruff, T. A. Delchar, *Modern Techniques of Surface Science*, 2nd ed., Cambridge University Press, Cambridge **1994,** Chapter 3.
[20] N. R. Campbell, D. Ritchie, *Photoelectric Cells*, Pitman, London **1929,** p. 4.

[21] A. H. Sommer, *Photoemissive Materials*, Krieger, Huntington, New York **1980,** Chapter 10.
[22] D. P. Woodruff, A. M. Bradshaw, *Rep. Prog. Phys.* **1994,** *57*, 1029.
[23] E. Brüche, *Z. Phys.* **1933,** *86*, 448.
[24] C. Ramsauer, *Elektronenmikroskopie*, Springer, Berlin **1943,** p. 32.
[25] E. Brüche, H. Johannson, *Naturwissenschaften* **1932,** *20*, 353.
[26] M. Knoll, E. Ruska, *Z. Phys.* **1932,** *78*, 318.
[27] J. Pohl, *Physik. Z.* **1934,** *35*, 1003.
[28] H. Mahl, J. Pohl, *Z. Tech. Phys.* **1935,** *16*, 219.
[29] H. Gross, G. Seitz, *Z. Physik* **1937,** *105*, 734.
[30] G. Möllenstedt, F. Lenz, *Advances in Electronics*, Vol. 13 **1963,** p. 251.
[31] L. Wegmann, *J. Microsc.* **1972,** *96*, 1.
[32] R. A. Schwarzer, *Microsc. Acta* **1981,** *84*, 51.
[33] O. H. Griffith, G. F. Rempfer, *Advances in Optical and Electron Microscopy*, Vol. 10 **1987,** p. 269.
[34] E. Bauer, W. Telieps in *Surface and Interface Characterization by Electron Optical Methods* (Eds.: A. Howie and U. Valdré), Plenum, New York **1988,** p. 195.
[35] O. H. Griffith, W. Engel, *Ultramicroscopy* **1991,** *36*, 1.
[36] M. Mundschau, *Synchrotron Radiation News* **1991,** *4*, 29.
[37] M. E. Kordesch in *CRC Handbook of Surface Imaging and Visualization* (Ed.: A. T. Hubbard) CRC Press, Boca Raton **1995,** Chapter 56.
[38] G. P. Harnwell, J. J. Livingood, *Experimental Atomic Physics*, McGraw-Hill, New York **1933,** 279.
[39] W. Koch, *Z. Phys.* **1958,** *152*, 1.
[40] L. Wegman, *Z. Angew. Phys.* **1969,** *27*, 199.
[41] J. G. Eden, *Photochemical Vapor Deposition*, Wiley, New York **1992,** p. 59.
[42] G. Margaritondo, *Introduction to Synchrotron Radiation*, Oxford University Press, Oxford **1988.**
[43] B. P. Tonner, G. R. Harp, *Rev. Sci. Intrum.* **1988,** *59*, 853.
[44] B. P. Tonner, G. R. Harp, *J. Vac. Sci. Technol. A* **1989,** *7*, 1.
[45] B. P. Tonner, *Nucl. Instr. & Methods A* **1990,** *291*, 60.
[46] F. Polack, S. Lowenthal in *X-ray Microscopy* (Eds.: G. Schmahl, D. Rudolph), Springer, Berlin **1984,** p. 251.
[47] L. Y. Huang, *Z. Phys.* **1957,** *149*, 225.
[48] M. D. Jones, G. A. Massey, D. L. Habliston, O. H. Griffith, *Beitr. elektronenmikroskop. Direktabb. Oberfl.* **1979,** *12*(2), 177.
[49] E. Bauer, *Rep. Prog. Phys.* **1994,** *57*, 895.
[50] L. H. Veneklasen, *Rev. Sci. Instrum.* **1992,** *63*, 5513.
[51] C. E. Hall, *Introduction to Electron Microscopy*, McGraw-Hill, New York **1953.**
[52] V. E. Cosslet, *Introduction to Electron Optics*, 2nd ed., Oxford University Press, Oxford **1950.**
[53] O. Klemperer, *Electron Optics*, 2nd ed., Cambridge University Press, Cambridge **1953.**
[54] B. Paszkowski, *Electron Optics*, American Elsevier, New York **1968,** 199.
[55] E. F. Burton, W. H. Kohl, *The Electron Microscope*, 2nd ed., Reinhold, New York **1948.**
[56] R. A. Schwarzer, *Beitr. elektronenmikroskop. Direktabb. Oberfl.* **1979,** *12*(2), 3.
[57] Y. Uchikawa, S. Maruse, *J. Electron Microscopy* **1970,** *19*, 12.
[58] G. N. Plass, *J. Appl. Phys.* **1942,** *13*, 49.
[59] H. Liebl, *Optik* **1979,** *53*, 1.
[60] V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier, A. W. Vance, *Electron Optics and the Electron Microscope*, Wiley, New York **1945.**
[61] M. Mundschau, E. Bauer, W. Swiech, *Surf. Sci.* **1988,** *203*, 412.
[62] G. F. Rempfer, W. P. Skoczylas, O. H. Griffith, *Ultramicroscopy* **1991,** *36*, 196.
[63] M. Mundschau, E. Bauer, W. Telieps, W. Swiech, *Surf. Sci.* **1989,** *213*, 381.
[64] E. Bauer, M. Mundschau, W. Swiech, W. Telieps, *Ultramicroscopy* **1989,** *31*, 49.
[65] M. Mundschau, E. Bauer, W. Swiech, *J. Appl. Phys.* **1989,** *65*, 581.
[66] M. Mundschau, E. Bauer, W. Swiech, *Phil. Mag. A* **1989,** *59*, 217.
[67] M. Mundschau, *Eur. Microsc. & Analysis* **1990,** (Sept.) 17.
[68] M. Mundschau, E. Bauer, W. Swiech, *Metall. Trans. A* **1991,** *22A*, 1311.
[69] M. Mundschau, E. Bauer, W. Swiech, *Mat. Res. Soc. Symp. Proc.* **1993,** *280*, 333.
[70] W. K. Burton, N. Cabrera, F. C. Frank, *Phil. Trans. R. Soc. London* **1951,** A*243*, 299.
[71] E. Bauer, M. Mundschau, W. Swiech, W. Telieps, *Inst. Phys. Conf. Ser.* **1988,** Vol. 1, No. 93, p. 213.
[72] E. Bauer, M. Mundschau, W. Swiech, W. Telieps in *Evaluation of Advanced Semiconductor Materials by Electron Miscoscopy* (Ed.: D. Cherns), Plenum, New York **1990,** p. 283.
[73] M. Mundschau, E. Bauer, W. Telieps, W. Swiech, *J. Appl. Phys.* **1989,** *65*, 4747.
[74] S. K. Scott, *Oscillations, Waves and Chaos in Chemical Kinetics*, Oxford University Press, Oxford **1994.**
[75] S. K. Scott, *Chemical Chaos*, Oxford University Press, Oxford **1993.**
[76] P. Gray, S. K. Scott, *Chemical Oscillations and Instabilities*, Oxford University Press, Oxford **1990.**

[77] P. Grindrod, *Patterns and Waves*, Oxford University Press, Oxford **1991**.
[78] V. A. Vasilev, Yu. M. Romanovski, D. S. Chernavskii, V. G. Yakhno, *Autowave Processes in Kinetic Systems*, Deutscher Verlag der Wissenschaften, Berlin **1987**.
[79] H. H. Rotermund, W. Engel, S. Jakubith, A. von Oertzen, G. Ertl, *Ultramicroscopy* **1991,** *36*, 164.
[80] H. H. Rotermund, W. Engel, M. Kordesch, G. Ertl, *Nature* **1990,** *343*, 355.
[81] G. Ertl, *Science* **1991,** *254*, 1750.
[82] S. Jakubith, H. H. Rotermund, W. Engel, A. von Oertzen, G. Ertl, *Phys. Rev. Lett.* **1990,** *65*, 3013.
[83] H. H. Rotermund, S. Jakubith, A. von Oertzen, G. Ertl, *Phys. Rev. Lett.* **1991,** *66*, 3083.
[84] M. Bär, S. Nettesheim, H. H. Rotermund, M. Eiswirth, G. Ertl, *Phys. Rev. Lett.* **1995,** *74*, 1246.
[85] M. D. Graham, I. G. Kevrekidis, K. Asakura, J. Lauterbach, K. Krischer, H. H. Rotermund, G. Ertl, *Science* **1994,** *264*, 80.
[86] M. Mundschau, M. E. Kordesch, B. Rausenberger, W. Engel, A. M. Bradshaw, E. Zeitler, *Surface Sci.* **1990,** *227*, 246.
[87] M. Mundschau, B. Rausenberger, *Platinum Metals Rev.* **1991,** *35*, 188.
[88] W. Engel, M. E. Kordesch, H. H. Rotermund, S. Kubala, A. von Oertzen, *Ultramicroscopy* **1991,** *36*, 148.
[89] G. V. Spivak, T. N. Dombrovskaia, *Soviet Physics-Doklady* **1956,** *106*, 9.
[90] G. V. Spivak, T. N. Dombrovskaia, N. N. Sedov, *Soviet Physics-Doklady* **1957,** *107*, 120.
[91] K. Schur, G. Pfefferkorn, *Beitr. elektronenmikroskop. Direktabb. Oberfl.* **1971,** *4*(2), 235.
[92] M. Mundschau, J. Romanowicz, *J. Vac. Sci. Technol. A*, to be published.
[93] B. P. Tonner, D. Dunham, J. Zhang, W. L. O'Brien, M. Samant, D. Weller, B. D. Hermsmeier, J. Stöhr, *Nucl. Instr. & Methods A* **1994,** *347*, 142.
[94] J. Stöhr, Y. Wu, B. D. Hermsmeier, G. R. Harp, B. P. Tonner, *Science* **1993,** *259*, 658.
[95] L. Weber, H. R. Oswald, E. Grauer-Carstensen, *Micron* **1975,** *6*, 129.
[96] M. E. Kordesch, *J. Vac. Sci. Technol. A*, **1995,** *13*, 1517.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

2 Field Emission and Field Ion Microscopy (Including Atom Probe FIM)

2.1 Field Emission Microscopy

The field emission microscope (FEM) is a projection microscope, in which an enlarged image of a sharply pointed object is produced on a fluorescent screen (Fig. 1). The image is produced in vacuum, by applying a strong negative potential to the specimen, which must be an electrical conductor. At sufficiently high local field strengths, electrons can escape from the specimen by tunneling through the deformed surface potential barrier (Fig. 2). The intensity of the field emission current I is a function of the local field strength F and the work function of the surface ϕ, and is expressed mathematically by the Fowler–Nordheim equation:

$$\frac{I}{F^2} = a \exp\left(-\frac{b\phi^{\frac{3}{2}}}{F}\right) \tag{1}$$

where a and b are effectively constants.

For electron tunneling to occur at a measurable rate, field strengths of 1–5 V nm^{-1} are necessary. In order to obtain such fields with the application of only a few thousand volts, a sharp pointed geometry for the specimen is essential. For the case of an isolated charged sphere, the field strength at the surface is given by:

$$F = \frac{V_0}{r} \tag{2}$$

where V_0 is the applied voltage, and r is the radius of curvature. For a needle-like specimen, which can be approximated as a sphere on a cone, the apex field strength may be expressed as:

$$F = \frac{V_0}{kr} \tag{3}$$

where k is a constant depending on the half-angle of the conical shank, and is approximately equal to 5 for typical field emission specimens.

The trajectories of the electrons leaving the surface are approximately radial. Hence the magnification of the image is given by:

$$M = \frac{D}{\beta r} \tag{4}$$

where D is the specimen-to-screen distance, and β is an image compression factor (approximately 1.5) which takes into account the effects of the specimen shank. Thus, for specimens of end radii below 0.5 μm, and specimen-to-screen distances of the order of 100 mm, magnifications of the order of 100 000 are obtained.

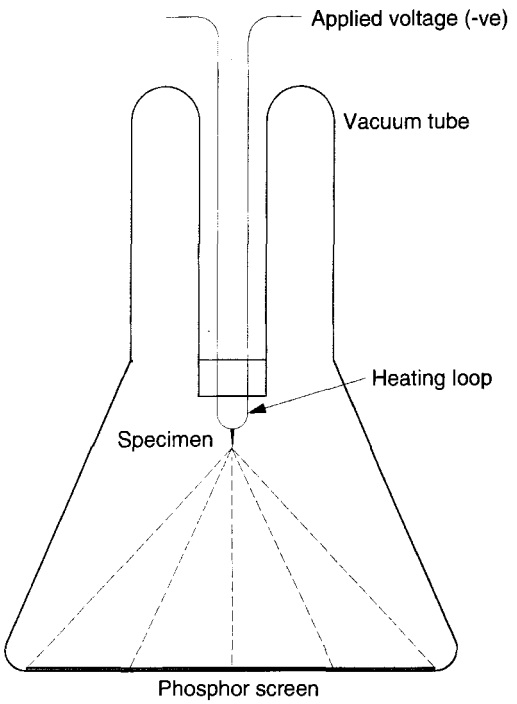


Figure 1. Schematic diagram of an FEM. A needle-like specimen is mounted along the axis of a vacuum tube, across the end of which is situated a fluorescent screen. Application of a high negative voltage to the specimen results in electron emission, and the production of a greatly enlarged image of the apex region on the microscope screen.

The resolution of the FEM image is limited by two main factors, the finite wavelengths of the emitted electrons, and the components of their momentum transverse to the emitting surface. The resolutions can be expressed by an equation of the form [1]:

$$\delta = C_1 \sqrt{r} \sqrt{\left(\frac{C_2}{\sqrt{V}} + \frac{C_3}{\sqrt{\phi}}\right)} \qquad (5)$$

where C_1, C_2, and C_3 are constants. The first term in the brackets is related to the electron wavelength, and the second term arises from the distribution of transverse momentum components of the emitted

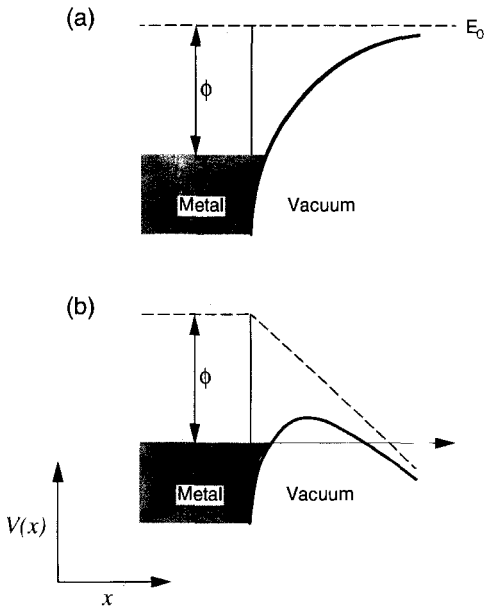


Figure 2. One-dimensional potential energy diagram for an electron on a metal surface: (a) under field-free conditions; (b) in the presence of a strong negative field. The application of the field produces a narrow energy barrier, through which tunneling of electrons can occur, resulting in field emission.

electrons. Equation (5) shows that δ is independent of the specimen-to-screen distance, and varies with the square root of tip radius.

It should be noted that the electrons which are emitted are predominantly from energy levels at or near the Fermi surface. This is because these particles face the smallest barrier to escape (Fig. 2). Because of the quantized nature of the electron energy levels in solids, these electrons have inherently rather high energies (corresponding, in Maxwell–Boltzmann terms, to temperatures of many thousands of degrees kelvin). Hence, transverse momentum components tend to be high, and the point-to-point resolution of the FEM is limited to around 2 nm. It is not possible to improve on this resolution limit by, for

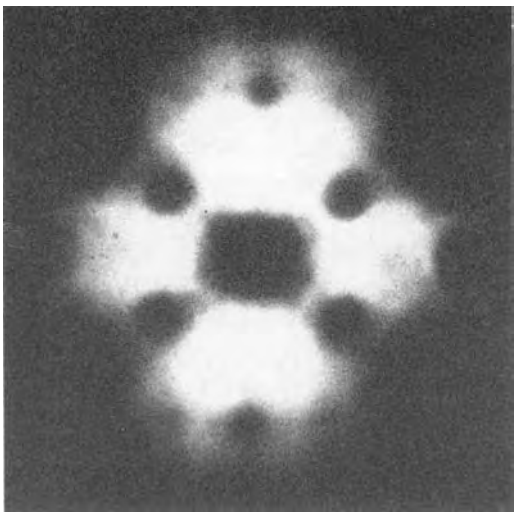


Figure 3. FEM image of a clean, thermally annealed tungsten specimen. The (011) plane is at the center.

example, cooling the specimen, because the Pauli exclusion principle prevents electron transitions to lower energy states.

Contrast in FEM images may arise from local variations in work function, field strength, or (very commonly) a combination of both. For example, Fig. 3 shows an FEM image of a single crystal tungsten specimen, prepared from tungsten wire by an electropolishing process, and then thermally cleaned by 'flashing' to a temperature in excess of 2000 K. The heating operation produces a rounded, almost hemispherical end-form on the specimen, with flat facets corresponding to low-index crystallographic planes such as {011}, {002}, and {112}. Atomically flat, low-index planes tend to have high work functions, and thus appear dark in the image, whereas the atomically rough, high-index planes have lower work functions and generally appear brighter.

Because of the radial nature of electron trajectories in the FEM, the symmetry of the specimen is preserved in the image, which resembles a stereographic projection of the original surface. Hence crystallographic indexing can usually be carried out very straightforwardly, by reference to standard projection maps. Figure 4 shows a [011] oriented projection corresponding to the orientation of the specimen in Fig. 3.

The main applications of the FEM have been to the study of adsorption and surface diffusion on metal and semiconductor surfaces, the investigation of thin film nucleation and growth processes, and the study of surface reactions such as oxidation and carburization. Field emitters are also used as electron sources of high intensity and good coherence. These have found wide application in other forms of microscopy, such as scanning electron (SEM), transmission electron (TEM), and scanning transmission electron microscopies (STEM), and electron holography. The main limitations of the field emission microscopy technique are its limited spatial resolution, and the absence of any simple means of chemical analysis of the specimen surface. These limitations were largely overcome by the introduction of the field ion microscope and the atom probe mass spectrometer, which are described below.

2.2 Field Ion Microscopy

The field ion microscope (FIM) allows the individual atoms on the surface of a solid to be imaged. It also permits the study of the three-dimensional structure of a material. Successive atom layers can be stripped from the surface in the form of ions, by the process of field evaporation. The material

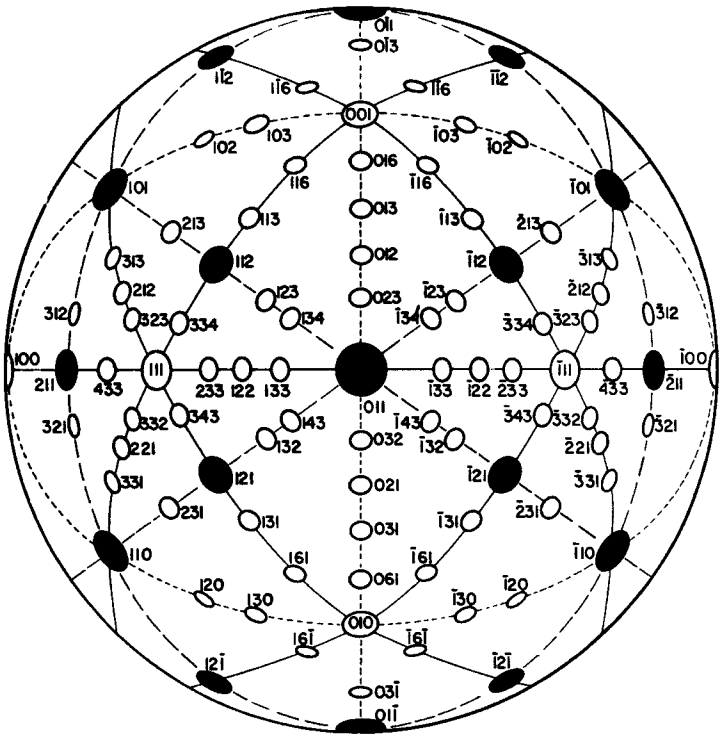


Figure 4. [011] centered projection of a cubic crystal structure. The orientation corresponds to the micrograph shown in Fig. 3. (Courtesy of Academic Press.)

removed from the surface by field evaporation can be chemically identified by linking the microscope to a time-of-flight mass spectrometer of single-particle sensitivity. This is known as an atom probe (AP). Nearly all metals and semiconductors can be examined by FIM/AP methods, but insulators and mechanically soft substances such as organic and biological materials are not amenable to study with this technique.

2.2.1 Principle of the Field Ion Microscope

The basic design of the FIM (Fig. 5) is similar to that of the FEM. There are some important differences, however. The needle-like specimen is usually made sharper (50–100 nm end radius), and is kept at low temperature (about 20–70 K). A high positive voltage (5–20 kV) is applied, rather than a negative one. Also, a small

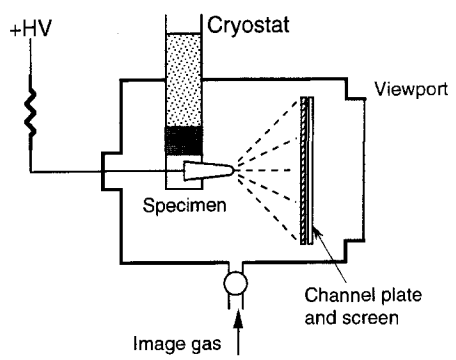


Figure 5. Schematic diagram of an FIM. Note that the specimen, which is typically 10 mm long and has an end radius of 100 nm, is not shown to scale.

amount of inert gas, such as helium or neon, is admitted to the microscope vacuum chamber. If the field is high enough, gas atoms which pass close to the specimen apex can become ionized. The resulting positive ions are accelerated away from the specimen along approximately radial paths, and strike a microchannel plate (MCP) image converter assembly. The MCP converts the incident ion beams to intensified cascades of secondary electrons, which are then accelerated towards a phosphor screen, where the final image is produced. As in the FEM, the radial projection of the ions produces a very high magnification of the image. The use of ions, rather than electrons, to form the image effectively removes the diffraction limit on the spatial resolution. Also, the energies of gas atoms are not subject to the same quantization constraints as for electrons in solids, so cooling the specimen is beneficial in reducing the transverse momentum components of the imaging particles. The result is that the FIM has a point-to-point spatial resolution approaching 0.2 nm, which is sufficient to image individual atoms on the specimen surface.

2.2.2 Field Ionization

The field ionization process involves quantum mechanical tunneling of an electron from the upper level of the image gas atom, into a vacant energy level in the specimen. Ground state ionization energies for the inert gases typically used for field ion imaging are high (10–25 eV). The applied electric field distorts the potential well around the atom, as shown in Fig. 6.

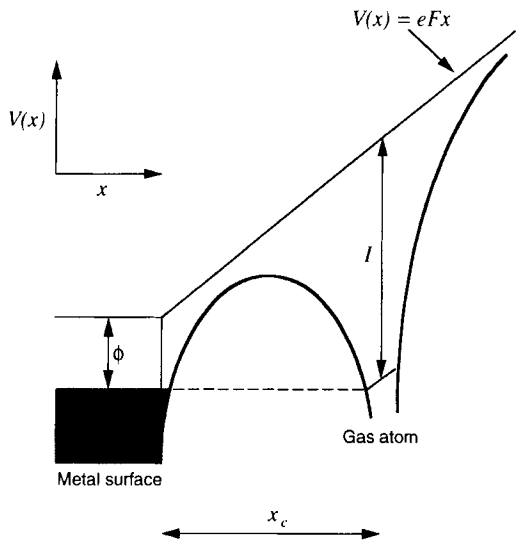


Figure 6. Electron energy diagram for a gas atom above a metal surface under the influence of a large electric field. The field distortion of the potential well associated with the gas atom produces a barrier through which the uppermost electron in the gas atom can tunnel into an empty state in the metal. At sufficiently high fields, the barrier is thin enough for a significant ionization probability.

A narrow (approximately 0.5 nm wide) potential barrier is thus formed, through which electron tunneling can occur, resulting in ionization of the gas atom. Two conditions must be satisfied for there to be a significant probability of electron tunneling. The first condition is that the gas atom must be sufficiently distant from the specimen surface for the energy level in the gas atom to be above the Fermi level in the specimen, so that an empty final state exists. The critical distance x_c at which ionization can take place is given approximately by:

$$x_c = \frac{I - \phi}{eF} \tag{6}$$

where I is the ionization energy of the gas atom, ϕ is the work function of the metallic specimen (more generally the Fermi

Table 1. Ionization potentials and imaging fields for selected gases (after Müller and Tsong [2])

Gas	Ionization potential (eV)	Best imaging field ($V\,nm^{-1}$)
Helium	24.6	44
Neon	21.6	35
Hydrogen	15.4	22
Argon	15.8	22
Krypton	14.0	15
Xenon	12.1	12

energy), F is the applied field, and e is the electronic charge. The second condition is that the potential barrier must be narrow enough to permit tunneling. In order for this to occur, the gas atom must be close to the specimen surface; in general, the probability for tunneling falls rapidly at distances beyond about 0.5 nm. These two conditions can only be met with extremely strong fields, above $10^{10}\,V\,m^{-1}$, that is at fields approaching an order of magnitude higher than those needed for field electron emission. A specimen with an apex radius of 50 nm will require voltages in the range 5–10 kV to attain the required ionization fields of common image gases such as neon or helium. Table 1 shows the ionization fields required for a number of gases used in field ion microscopy.

2.2.3 Field Evaporation

Under the intense fields produced in the FIM, atoms of the specimen material can also become ionized, and this process is termed field evaporation. It is usually a thermally activated process, except at the temperatures below approximately 50 K, where it is believed that ion tunneling may occur. The field reduces the energy barrier for desorption to the point that removal of material occurs even at cryogenic temperatures. Potential energy diagrams for a neutral atom and a positive ion at the surface of an FIM specimen are shown schematically in Fig. 7. Under zero-field conditions, the energy $Q_{0(n)}$ required to remove a surface atom as an n-fold charged ion is given by:

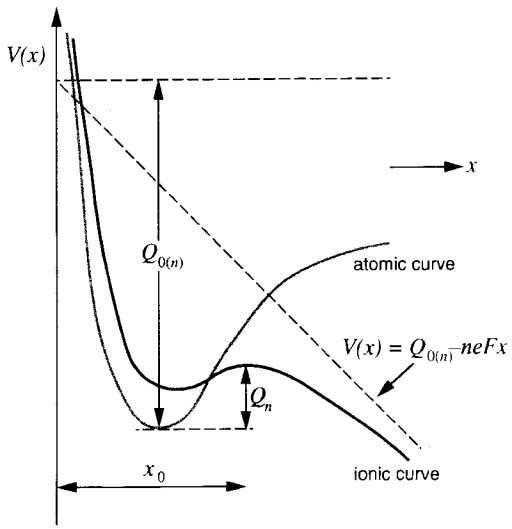


Figure 7. Schematic potential energy diagrams for an atom and an ion above the surface of a metal specimen in the presence of a large electric field. The field lowers the ion energy, so that at some distance from the specimen surface it is energetically favorable for ionization to occur. This crossover leads to a reduced activation barrier (energy Q_n) for evaporation of a surface atom.

$$Q_{0(n)} = \Lambda + \sum_{k=1}^{n} I_k - n\phi \qquad (7)$$

where Λ is the sublimation energy, I_k is the kth ionization energy of the atom, and ϕ is the work function of the material. The presence of the electric field F at the specimen surface reduces the potential energy of the ion at a distance x from the surface by $neFx$, so that the ionic and atomic curves intersect, as shown in Fig. 7. In

the simplest model of field evaporation, repulsion terms in the ionic potential are ignored, and the potential energy of the ion in the region of the crossover is expressed approximately by the sum of an image potential and a field term:

$$V(x) = Q_{0(n)} - \frac{n^2e^2}{16\pi\epsilon_0 x} - neFx \qquad (8)$$

where ϵ_0 is the permittivity of free space. This potential curve has a maximum defined by $dV/dx = 0$ which occurs at a distance x_0 from the specimen surface

$$x_0 = \frac{1}{4}\sqrt{\frac{ne}{\pi\epsilon_0 F}} \qquad (9)$$

Assuming that the neutral and ionic potential energy curves cross closer to the specimen than x_0, the maximum in the ionic potential (the 'image hump') represents the energy barrier to the removal of a surface atom as an n-fold charged positive ion. The activation energy Q_n is therefore given by:

$$Q_n = Q_{0(n)} - \sqrt{\frac{n^3e^3F}{4\pi\epsilon_0}} \qquad (10)$$

Approximate fields required for evaporation at cryogenic temperatures are readily calculated by setting $Q_n = 0$. For any given element, the fields calculated for the evaporation of singly, doubly, and triply charged ions will generally be different. The evaporation field of that material is taken to be the lowest of these values. Evaporation fields calculated using this method are shown in Table 2 for a number of elements, together with the predicted charge states of the ions generated. Despite the simplicity of the model, the agreement with the experiment is remarkably good. For a fuller treatment of this and other models of field evaporation, the reader is referred to Tsong [3] or Miller et al. [4].

2.2.4 Image Formation, Magnification, and Resolution

The field ion image is produced by a combination of several high-field effects, as indicated schematically in Fig. 8. Atoms of the image gas which arrive in the apex region are polarized and attracted towards the specimen. Their initial collision with the surface results in partial thermal accommodation to the temperature of the specimen, so that the atom is trapped at the surface. The atom executes a form of hopping motion across the surface, and multiple impacts lead to further thermal accommodation. The supply of gas atoms directly to the specimen apex, as shown in Fig. 8, is enhanced by gas collected on the shank, which diffuses preferentially towards the apex under the influence of the applied field. Some of the gas atoms which diffuse over the apex of the specimen become bound at specific locations on the surface, principally those at which the field is highest, such as kink site atoms. This field adsorbed layer increases the efficiency of thermal accommodation for other gas atoms incident on the surface. Gas atoms which pass over one of these high field sites may be field ionized, and then travel away from the specimen towards the image screen, to build up the field ion image.

The FIM image (Fig. 9) shows individual atoms on the specimen surface as distinct, well-resolved spots, arranged in patterns of concentric circles. As the process of field ionization is strongly field

Table 2. Calculated and experimental evaporation fields (after Müller and Tsong [2] and Tsong [5])

Material	Calculated field for most easily evaporated species ($V\,nm^{-1}$)			Observed field ($V\,nm^{-1}$)	Main observed species
	M^+	M^{2+}	M^{3+}		
Be		46		34	Be^{2+}, Be^+
Al	19			33	Al^+, Al^{2+}
Si		32		30	Si^{2+}
Ti		26		25	Ti^{2+}
V		30			V^{2+}
Cr	27	29			Cr^{2+}
Mn	30	30			Mn^{2+}
Fe		33		35	Fe^{2+}
Co		37		36	Co^{2+}
Ni	35	36		35	Ni^{2+}
Cu	30			30	Cu^+
Zn	33				Zn^{2+}
Ga	15				
As		42			
Zr		28		35	Zr^{2+}
Nb		37		35	Nb^{2+}
Mo		41		46	Mo^{2+}, Mo^{3+}
Ru		41		45	
Rh		41		46	Rh^{2+}, Rh^{3+}
Pd	37				Pd^+
Ag	24				Ag^+
In	13				
Sn		23			
Hf		39			Hf^{3+}
Ta			44		Ta^{3+}
W			52	57	W^{3+}
Re		45		48	Re^{3+}
Ir		44		53	Ir^{3+}, Ir^{2+}
Pt		45		48	Pt^{2+}
Au	53	54		35	Au^+, Au^{2+}

dependent, gas atoms are ionized preferentially above the highest field sites, which tend to be corner or kink site atoms. Each set of concentric rings reveals the prominent atoms surrounding an individual crystallographic pole, with the directions of largest interplanar spacing being the most prominent. It should be noted that the image rings are not due to a diffraction effect, but are a direct, real-space image, arising from the hemispherical shape of the specimen apex. The ball model shown in Fig. 10 has a large number of spheres, arranged on a face-centered cubic lattice and shaped into an approximately hemispherical end-form. The spheres on kink sites, shown in white, form a pattern similar to the field ion image.

An important concept in field ion microscopy is the best image field (BIF). This refers to the field strength at which the best overall image contrast is obtained for a given specimen, in a given image gas. If the applied voltage is too low, the field is inadequate to produce an ion current from the specimen, while if the field is too high

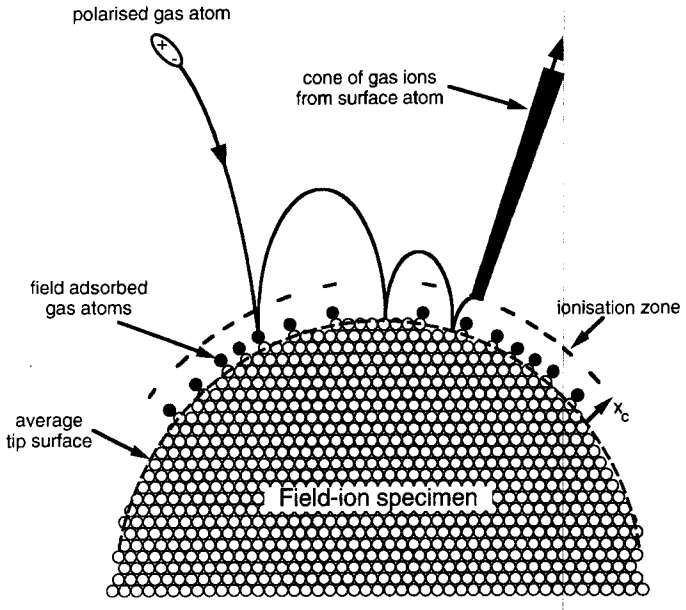


Figure 8. Schematic diagram showing the processes which occur in the region of the specimen apex which lead to the formation of a field-ion image. Polarization forces attract gas atoms towards the specimen surface where they are thermally accommodated to the specimen temperature through a series of collisions. Some of these atoms are trapped at specific sites on the specimen surface, while others become field ionized above prominent atom sites and contribute to the image.

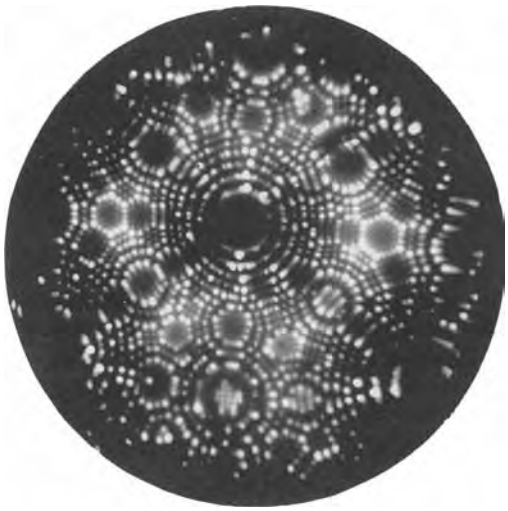


Figure 9. Field ion micrograph from a tungsten single crystal. The image was recorded at a specimen temperature of 28 K, using helium as the image gas, and with an applied voltage of 4.5 kV. The orientation of the sample was the same as that for the field emission micrograph shown in Fig. 3 and the crystallographic projection in Fig. 4.

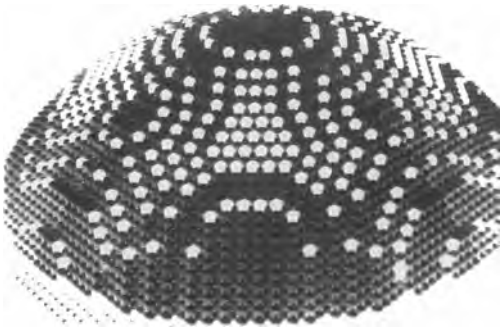


Figure 10. Ball model of the specimen apex, constructed from a number of spheres arranged in a cubic lattice. The prominent sites, representing kink site atoms in the surface of the specimen, are shown white. These form intersecting sets of concentric rings similar to those seen in the field ion image.

ionization occurs uniformly above the tip surface and the image becomes fogged out. Image contrast is due to fluctuations in the overall rate of field ionization between different points on the specimen surface. There has been much debate about whether this is mainly associated with variations in the local concentration of image gas atoms

on the surface, or to variations in tunneling probability. Both effects would be expected to vary with local field strength and surface charge, and it now seems that both effects are important.

Additional contrast effects arise due to the phenomenon of field adsorption. Field ionization may take place by tunneling through an adsorbed gas atom, rather than through a vacuum barrier. There is no complete theory of this process, but field adsorption binding energies are of the order of 10–20 kJ mol^{-1}, and so the surface will be almost completely covered with adsorbed gas atoms at the lowest imaging temperatures. The presence of adsorbed gas atoms does not seem to affect the determination of atom positions in the FIM, since the adsorbed atoms are located directly above the most prominent substrate atoms. However, if neon or argon atoms are adsorbed on the surface, the brightness of the image is increased, and the BIF is decreased, since the rate of electron tunneling through such atoms is greater than through helium atoms or a vacuum barrier. Small additions of hydrogen are also effective in lowering the BIF when helium or neon are used as the primary image gas, and this effect is also believed to be due to field adsorption. However, the addition of hydrogen commonly leads to the formation of metal hydride ions during field evaporation, and this leads to difficulties when carrying out atom probe microanalysis (see below).

As in the case of the FEM, the high magnification obtained in the FIM is due to the near-radial projection of ions from the specimen. With an apex radius of 50 nm, and a screen placed 10 cm away, a magnification of the order of 10^6 is achieved. The inverse dependence of magnification on apex radius means there will be significant variations between specimens, and it is therefore important to record the applied voltage at which each FIM image is obtained. The spatial resolution in the FIM is limited at the lowest temperatures by the size of the ionization 'disk' above high field sites on the specimen. At higher operating temperatures (>50 K), the resolution is controlled by the thermal velocities of the gas atoms at the moment of ionization. The lateral component of velocity leads to an angular spread in the trajectories of ions emitted from any given site and a corresponding increase in the size of the image spot. For a specimen temperature T this gives an approximate resolution for the microscope of:

$$\delta = 4r\beta\sqrt{\frac{k_B T}{eV_0}} \tag{11}$$

For a specimen radius of 100 nm and an applied voltage of 10 kV, and taking β to be 1.5 gives a resolution of 1 nm at 300 K but 0.24 nm at 20 K.

A distinction should be made between the resolution of the image as defined by the ability to separate two atoms, and the precision with which the position of a single atom may be located. By determining the center of each spot in a well-resolved region of a field ion image, it is sometimes possible to measure atomic positions to within 0.02 nm.

2.2.5 Contrast from Lattice Defects and Alloys, and Analysis of Field Ion Microscope Images

If the apex of the specimen contains a grain boundary, the corresponding field

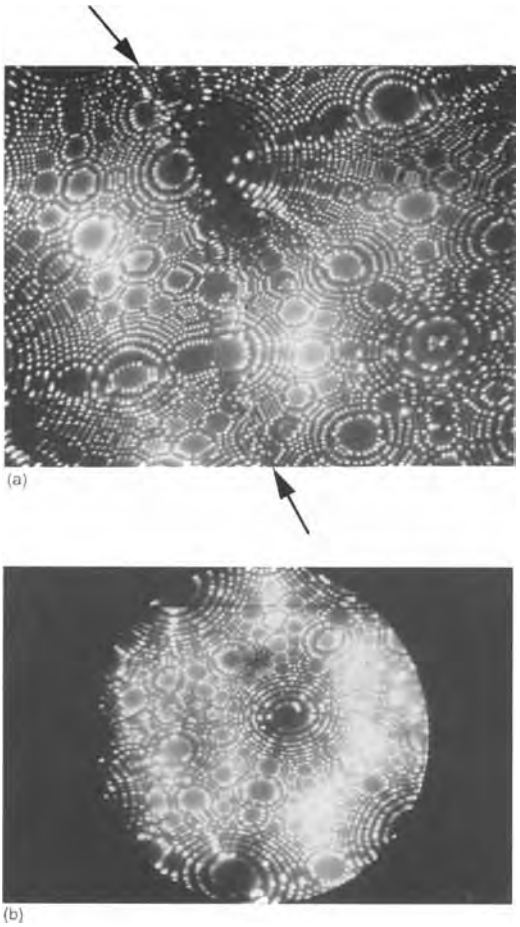


Figure 11. (a) Field ion micrograph of a tungsten specimen with a grain boundary. The presence of the boundary is seen as a discontinuity in the regular crystallographic contrast of the individual grains. (b) Field ion micrograph of a tungsten specimen containing a dislocation. The intersection of the dislocation with the surface of the specimen in the central (011) region converts the usual ring-like structure into a spiral. In order to produce this form of contrast, the dislocation must have a Burgers vector component normal to the specimen surface at the point of intersection. (Courtesy of T. J. Godfrey.)

ion image exhibits rings which are discontinuous at the boundary (Fig. 11a). Other types of lattice defect can also be seen in the FIM, but are more difficult to observe. If a dislocation intersects the surface of the specimen close to a major crystallographic pole, and there is a component of the Burgers vector perpendicular to the surface, the image of the pole takes the form of a spiral (Fig. 11b). Point defects such as vacancies can also be observed, but because the field ion image does not show all the atoms on the surface, great care must be taken in the correct interpretation of such images.

When single-phase alloys are studied, there are often differences in the brightness of the field ion image of different atoms. This element-specific contrast may arise from preferential field ionization above different atoms, or from differences in the field evaporation behavior of the elements, or a combination of both effects. Species with high evaporation fields will be preferentially retained on the surface of the specimen, and will be associated with higher local fields and thus image more brightly. Elements with low evaporation fields will image darkly. In the case of concentrated solid solutions, the strong crystallographic contrast present in single-element materials is usually lost, because the image of the regular crystal structure is perturbed by the differences in image brightness from the constituent elements. In ordered materials in which there are large differences in the evaporation fields of the two species, exaggerated crystallographic contrast effects arise, in which only one of the elements is visible. Even when the difference in evaporation fields is not great, ordering on a superlattice pole can usually be observed as alternate bright and dim planes. Antiphase boundaries (APBs) are also visible in the field ion image as a mismatch between the bright/dark contrast of the individual atomic planes on either side of the boundary.

In the case of multiphase materials, there is often contrast which is very useful from a metallurgical point of view. This contrast results from differences in the local radius of curvature over the sample, due to differences in binding energy, and thus evaporation field, between the phases. A phase which has a lower evaporation field than the matrix must have a larger local radius of curvature than the surrounding material or it will field evaporate at a higher rate. Similarly, a region of high evaporation field must have a lower radius of curvature; that is, it must protrude slightly above the mean tip profile, or it will not evaporate sufficiently. Thus the shape which is produced by field evaporation is that which gives the same rate of removal from the whole imaged region. Less refractory regions will appear darker in the FIM image, because of reduced local fields, whereas more refractory phases will appear brighter. Examples of the images obtained from two-phase materials are shown in Fig. 12.

Once an FIM image has been established, field evaporation allows the internal structure of the material to be examined. The specimen is sectioned by removing the uppermost layers in a controllable manner. As the voltage is raised, atoms which occupy the most prominent positions on the surface (at edges and kink sites) are the first to field evaporate. The removal of these atoms exposes others to similarly high fields. The overall result is that every atomic terrace progressively shrinks in size as field evaporation progresses. The removal of single atomic layers can be observed from the collapse of individual rings in the images of low index poles, and thus the absolute depth traversed through the specimen can be measured. In this way, the microstructure can be observed in three dimensions. The material removed by the field evaporation process can be chemically identified by mass spectrometry, thus allowing the composition of the imaged regions to be determined (see below).

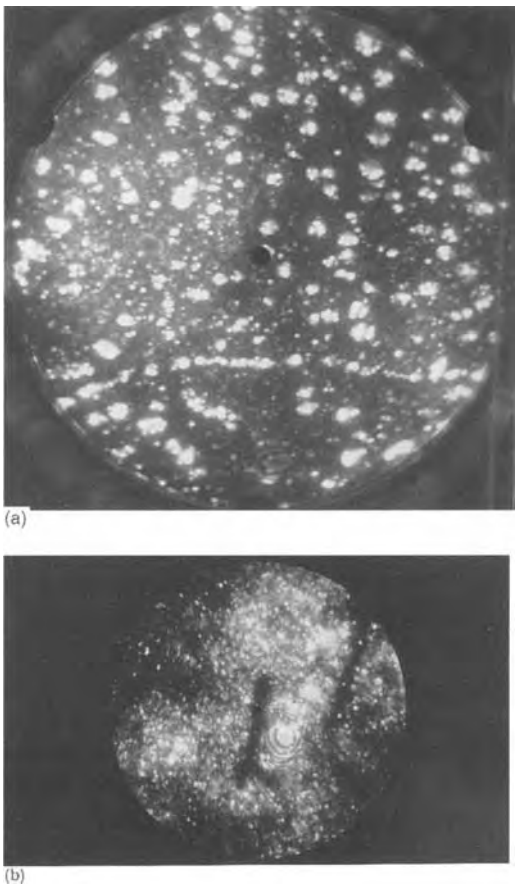


Figure 12. Examples of bright and dark contrast in field ion images, resulting from second-phase particles. (a) Image of a model maraging steel (Fe-18.2 at.% Ni-8.8% Co-2.9% Mo) specimen aged at 783 K for 1 h, showing brightly imaging spheroidal molybdenum rich precipitates. The distance across the image is approximately 165 nm. (Courtesy W. Sha and ASM International.) (b) Darkly imaging plates of cementite (Fe_3C) in a pearlitic steel (Fe-3.47 at.% C-0.65% Mn-0.6% Si). (Courtesy of M. K. Miller.)

Crystallographic parameters may be measured from FIM micrographs. As explained earlier, the image is a projection of the three-dimensional shape of the specimen surface. To a first approximation, this projection may be taken as stereographic. Major crystallographic poles may be indexed from the symmetry of the pattern, the main zone lines can then be constructed, and the indices of minor poles found from the zone addition rule. Once indexing has been carried out, the magnification of any given region of the image can be calculated. If the two-dimensional array of atoms on a given plane is fully resolved, and the lattice parameters and crystal structure of the material are known, then the magnification can be determined from direct measurement of the micrograph. More generally, if N is the number of image rings between a reference crystallographic pole (hkl) and another pole separated from it by an angle θ, then the radius of the specimen in this region is given by:

$$r(1 - \cos\theta) = Nd_{hkl} \qquad (12)$$

where d_{hkl} is the interplanar spacing for the (hkl) pole. The magnification can then be obtained from a calculation of the distance $r\theta$ between the two poles, and comparison with the equivalent distance on the micrograph.

Three-dimensional reconstruction of the specimen geometry is possible if an extended sequence of micrographs is recorded at intervals during field evaporation. However, full analysis is hindered by the fact that the specimen end-form is never truly hemispherical. One reason is that the work function (and hence the evaporation field) varies with crystallographic orientation. Accurate measurement of curvilinear distances across nonuniform specimen surfaces is difficult, except in very localized areas, and extensive computer calculation is often required. A simple measurement of distance may be made along the axis of the specimen if this coincides with a major crystallographic pole, since the step height of the atomic ledges of the axial pole will automatically provide a unit of distance. Long field evaporation sequences may be analyzed in this way, by recording the number of atom layers removed between successive micrographs. One application is to find the size of second-phase precipitate particles from their 'persistence depth', which is a more reliable estimate than lateral measurement based on an individual micrograph. Another important application is in the reconstruction of the atomic-scale topography of grain boundaries and interphase interfaces.

2.2.6 Specimen Preparation Techniques

The specimen needles required for the FIM (and FEM) may be prepared by chemical or electrolytic polishing, or by physical methods such as ion milling, depending on the nature and form of the starting material. It is convenient to start with materials of round cross-section and small diameter, such as wires. However, it is also possible to make samples from much larger objects (e.g. an aero engine turbine blade, a steel plate, or a semiconductor single crystal) by first slicing out a specimen blank of square cross-section, about 0.5 mm diameter and 10–15 mm long. Specimen blanks can also be prepared from very small objects, such as

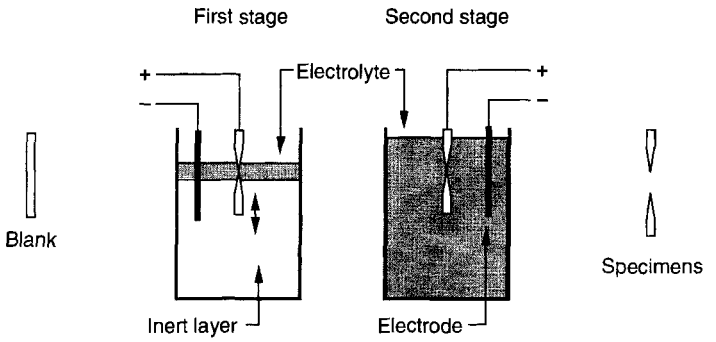


Figure 13. Method typically used for the preparation of field ion specimens from bulk metallic alloys. Specimen blanks are usually 10–15 mm long and with a width of approximately 0.5 mm.

evaporated thin films, using microlithographic techniques.

Electropolishing is the most widely used preparation technique for metals and alloys. A two-layer polishing method (Fig. 13) is often employed. A layer of electrolyte, about 5 mm deep, is floated on top of a heavier inert liquid, such as a chlorofluorocarbon fluid. As material is removed in the electrolyte layer, a neck develops on the specimen, until it fractures under its own weight. A small amount of further polishing usually results in a sharp specimen. Blunted samples can be resharpened by a micropolishing technique. The end of the needle is immersed in a single drop of electrolyte, suspended on a small platinum loop, 1–2 mm in diameter, which acts as the counter electrode. The micropolishing is carried out under an optical microscope, so that progress can be monitored continuously.

Chemical polishing is used mainly for poorly conducting materials such as semiconductors, for which electropolishing cannot be used. The method usually involves a repeated dipping technique, and great care is needed in order to produce a smooth specimen surface by this approach, as etching commonly occurs.

Ion milling is a slow but controllable method of FIM specimen preparation, and can be carried out in a conventional TEM specimen preparation unit, with some modification to the holder to accommodate the differently shaped specimen. The advantage of this method is that it generally produces a very smooth surface, without preferential etching at grain boundaries or second-phase particles, which can be a problem with the other techniques mentioned above. It is also suitable for specimens of low electrical conductivity.

A hybrid approach to specimen preparation which is useful in some circumstances is to begin the polishing using a chemical or electrolytic method, and then to use ion milling to carry out the final sharpening. A selection of recommended preparation techniques for different materials is given in Table 3.

2.3 Atom Probe Microanalysis

2.3.1 Principles of the Atom Probe Field Ion Microscope

The principle of the atom probe is outlined in Fig. 14. The simplest instrument consists

Table 3. Sample preparation techniques for selected materials

Material	Technique
Steels	First stage electropolish by two-layer method to form thin waist on specimen: 25% perchloric acid in glacial acetic acid, 20–25 V d.c., temperature 0–10°C. Final stage electropolish in a single bath of electrolyte: 2% perchloric acid in 2-butoxyethanol, 20–25 V d.c., temperature 0–10°C
Nickel alloys	First stage electropolish by two-layer method to form thin waist on specimen: 10% perchloric acid + 20% glycerol + 70% ethanol, 22 V d.c. Final stage electropolish in a single bath of electrolyte: 2% perchloric acid in 2-butoxyethanol, 25 V d.c.
Copper alloys	Electropolish in concentrated orthophosphoric acid, 16 V d.c.
Aluminum alloys	Procedures vary from alloy to alloy. One versatile method is to electropolish with 2–10% perchloric acid in methanol, 5–10 V a.c., at −10°C
Titanium alloys	Procedures vary from alloy to alloy. One method is to electropolish in 6% perchloric acid + 34% n-butyl alcohol + 60% methanol
Molybdenum alloys	First stage electropolish by two-layer method to form thin waist on specimen: 5 N aqueous sodium hydroxide solution, 6 V a.c. Final stage electropolish in a single bath of 12% concentrated sulphuric acid in methanol, 6 V d.c.
Platinum alloys	Electropolish in molten salt mixture of 80% sodium nitrate + 20% sodium chloride at 440–460°C, using a repeated dipping technique. Start at 5 V d.c., reducing to 3 V.
Tungsten	Electropolish using two-layer method in 5 wt.% aqueous sodium hydroxide solution, 5–6 V a.c.
Silicon	Chemically polish in a solution of 15% concentrated nitric acid, 80% hydrofluoric acid (40% solution) and 5% glacial acetic acid. Finish by ion milling if required
Gallium arsenide	Chemically polish in a solution of 44% concentrated sulphuric acid, 28% hydrogen peroxide (30% w/v solution) and 28% water, at about 60°C

Reproduced from Metals Handbook, 9th Edition, Volume 10, Materials Characterisation (1986), courtesy of ASM International.

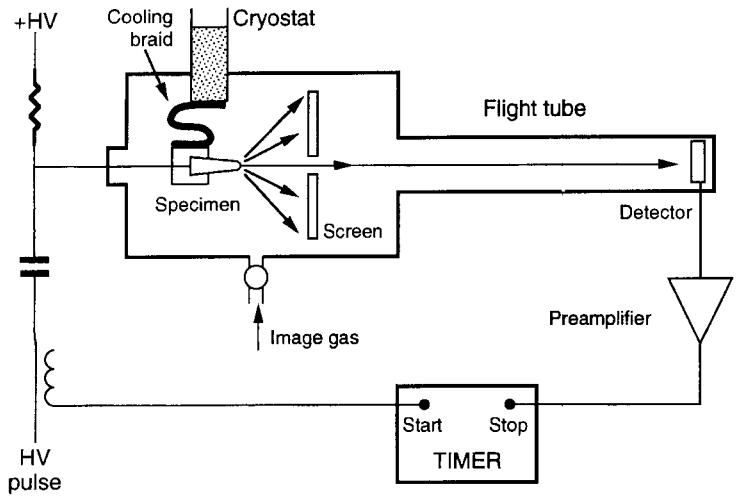


Figure 14. Schematic diagram of a straight flight path atom probe. The flight path is typically 1–2 m but can be as long as 8 m.

of a combination of an FIM and a linear time-of-flight mass spectrometer, connected via an aperture hole, about 2 mm diameter, located in the center of the microscope screen. The spectrometer consists of an evacuated flight tube, at the end of which is situated a single-ion sensitivity particle detector. The FIM specimen is mounted on a goniometer stage and can be tilted to align any part of the image over the aperture hole. Ions are field evaporated from the surface at well-defined moments in time by means of a succession of short high-voltage pulses, superimposed on the d.c. imaging voltage applied to the specimen. Typically, these pulses are of subnanosecond rise time, 10 ns duration, and have amplitudes of 15–20% of the d.c. voltage. The field evaporated ions are accelerated towards the microscope screen, following trajectories which are very closely similar to those of the gas ions which formed the original FIM image. Most of the ions are stopped by the screen, but those from the selected region of the specimen will pass through the aperture hole into the spectrometer. Because of the high magnification of the FIM, the 2 mm diameter aperture in the screen defines a region 1–2 nm in diameter on the specimen surface.

The flight time of the ions from the specimen to the detector is measured by equating the applied potential V_0 (the sum of the d.c. and the pulse voltages) with the final kinetic energy of the ion:

$$neV_0 = \frac{mv^2}{2} \tag{13}$$

where m is the mass of the ion and v its final velocity. The mass-to-charge ratio of the ion can be obtained from the flight time t over a length d:

$$\frac{m}{n} = \frac{2eV}{d^2}t^2 \tag{14}$$

This method does not allow the mass of an ion to be determined independently of its charge state, but fortunately the number of charge states produced by any given element is usually small. Under particular conditions of field and temperature, most elements will evaporate in only one or two states, as indicated in Table 2. The identification of mass peaks in the spectrum is therefore relatively straightforward. In certain cases, the analysis can be complicated by the overlap of elements with a common isobar (e.g. ^{58}Fe and ^{58}Ni) or by elements with different charge states (e.g. $^{56}Fe^{2+}$ and $^{28}Si^{+}$). When this occurs, reference data on natural isotopic abundances must be used to try and resolve the difficulty. These cases are few, however, and seldom pose a serious problem.

The lack of any spatially dispersive elements in the atom probe mass spectrometer design means it has a very high sensitivity, up to 100%, determined by the efficiency of the single-ion detector used in the instrument. The most common detectors are dual channel-plate assemblies (about 65% efficient). Channeltrons and multistage discrete dynode Cu–Be electron multipliers have also been used and, although less convenient to operate, give close to 100% efficiencies. However, the effectiveness of an atom probe is dependent not only on its sensitivity, but also on the accuracy with which it can identify a given ion. This is controlled by the spread of flight times for any given species, which in turn depends on the energy distribution of the field evaporated ions.

2.3.2 Energy Deficits and Energy Compensation

Because of the time varying electric field present at the specimen at the moment of evaporation, the ions are emitted with a small range of energies. This imposes a limit on the mass resolution of the simple linear atom probe of around $m/\Delta m = 200$ to 300 (full-width at half maximum). For the study of complex engineering alloys, this resolution limit can impose a serious restriction, especially in the case of steels where the mass-to-charge ratios of some of the important alloying elements (Mn, Co, Ni, Cu) lie in the 'tails' of the main iron peaks. To improve the mass resolution of the instrument, some form of energy compensation must be incorporated into the design. Modern atom probes now include this feature as standard. Two main approaches are used, either a toroidal sector lens in a configuration first published by Poschenrieder [6] or a reflectron lens, as developed by Karataev et al. [7]. An atom probe design incorporating a reflectron, which is an increasingly favored design, is shown in Fig. 15. Ions entering the reflectron will be turned back at different points, depending on their total energy.

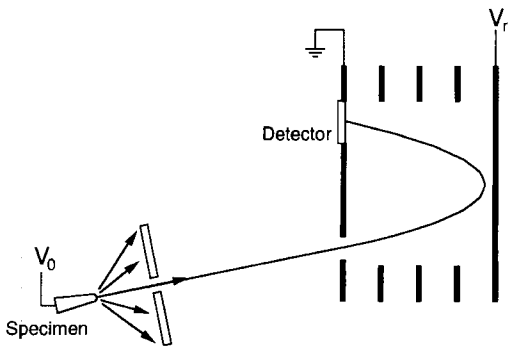


Figure 15. Schematic diagram of an energy compensated atom probe instrument using a reflectron. Ions of higher initial energy penetrate further into the reflectron. This increases their flight time inside the reflectron, so that they arrive simultaneously with ions of the same mass-to-charge ratio, but lower initial energy.

The highest energy ions will penetrate furthest into the reflectron, and will therefore have longer overall flight paths through the system. Thus, faster ions are effectively delayed, and the slower ions have an opportunity to 'catch up'. With suitable operating voltages, ions of the same mass-to-charge ratio can be time focused at the detector. Energy compensation is capable of producing an at least tenfold improvement in the mass resolution of the simple linear time-of-flight atom probe, making it possible to resolve individual isotopes fully (Fig. 16).

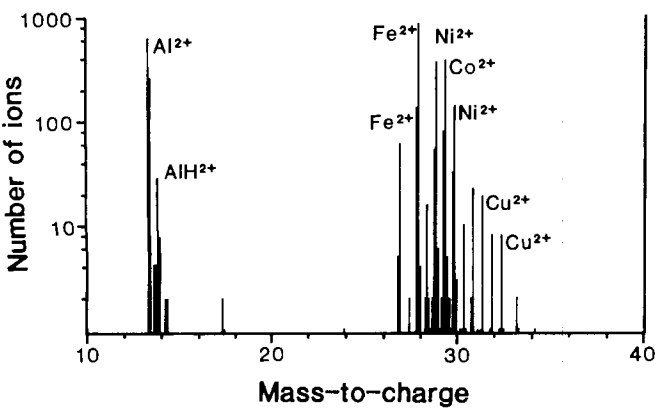


Figure 16. Energy compensated atom probe mass spectrum from the permanent magnet alloy, Alnico 5 (Fe-15.6 at.%Al-12.5%Ni-21.4%Co-2.5%Cu). At the high resolution available with this system, all the individual isotopes of the elements are well resolved.

2.3.3 Accuracy and Precision of Atom Probe Analysis

An essential factor in the operation of the atom probe is that the trajectories of field evaporated ions closely follow those of the corresponding image gas ions. Thus, when field evaporation occurs, only atoms whose images lie over the probe aperture pass into the time-of-flight mass spectrometer. The analysis area on the specimen surface is then defined by the region which corresponds to the physical aperture in the image screen. This limits the lateral resolution of conventional atom probe analysis. Some instruments allow the tip-to-screen distance to be altered, which allows the magnification of the FIM to be varied by a factor of 2–3. This has the secondary effect that the analysis area is also varied by the same amount, so that a smaller effective probe aperture can be selected if higher lateral resolution is required. However, as the analysis area becomes more restricted, the number of ions collected per atom plane field evaporated decreases. Therefore a compromise has to be achieved between spatial resolution and ion counting statistics.

In principle, the atom probe is a fully quantitative microanalytical tool. The method used to obtain an analysis is very straightforward. The absence of any need for external excitation to remove material from the specimen (e.g. by an electron beam or ion sputtering) means that interpretation of atom probe data is not affected by excitation cross-section, absorption, or fluorescence corrections. The specimen is held at cryogenic temperatures, and so atoms on the surface are only removed by field evaporation, and always in the form of positive ions. Accurate quantitative analysis only requires detection of the resulting ions with an efficiency which is independent of the atomic species. The efficiency of the channel plate detectors used in most atom probes is effectively independent of mass for ion energies of above 5 keV [8], and therefore creates no difficulties. The main limitation of the use of time-of-flight analysis is that ions are detected only if they are produced during the field evaporation pulse. Since the evaporation fields of the various elements present will generally be different, elements with lower evaporation fields may be evaporated preferentially at the d.c. voltage, leading to loss in sensitivity and also inaccuracy in the analysis. The ratio of pulse voltage to d.c. voltage must be kept sufficiently high that evaporation during the d.c. part of the cycle is negligible. Fortunately, the kinetics of field evaporation are strongly field dependent, with evaporation rates changing by up to an order of magnitude for only a 1% change in applied field, so pulse amplitudes of 15% of the d.c. standing voltage are usually sufficient. However, in some materials for which the evaporation fields of the constituent elements differ widely, higher pulse ratios (up to 20%) may be required. It is usually advisable to determine the range of conditions for quantitative analysis of a given type of alloy by performing reference calibration experiments on homogeneous, solution treated material of accurately known bulk composition, over a range of pulse ratios. As the field dependence of evaporation rate is highest at low temperatures, analysis should be performed at the lowest convenient temperature, in order to reduce the pulse amplitude required.

Once appropriate experimental conditions have been established, composition determination in the atom probe simply involves atom counting. However, since the volume of material analyzed in the atom probe is so small, the total number of atoms available for composition measurement is limited. The precision of the measurement is therefore limited by counting statistics. If there are n_s atoms of the species of interest in a total number of atoms N, the estimate for the concentration $c_0 = n_s/N$ has a standard deviation σ of:

$$\sigma = \sqrt{\frac{n_s(N - n_s)}{N^3}} \qquad (15)$$

For example, in an analysis of a steel in which a total of 10 000 atoms were detected, 2388 of these were Cr atoms. The composition from the analyzed volume is therefore estimated to be 23.8 ± 0.4 at.%.

Precise measurement of low atomic fractions in the atom probe is challenging, as it requires detection of very large numbers of atoms. When a trace element or impurity is localized within the material, for example at an interface, then the high spatial resolution of the atom probe allows precise measurement of the local concentration. The sensitivity of the ion detectors to light elements is a further advantage of the technique, allowing quantitative measurement of local concentrations of boron, carbon, or nitrogen, even in high atomic number matrix materials such as steels or nickel-based superalloys.

2.3.4 Atomic Plane Depth Profiling

The process of field evaporation provides depth profiling through the specimen. As analysis proceeds, data are collected from a cylinder of material approximately 2 nm in diameter, over a depth which depends on the total amount of material evaporated. The composition of this volume can be estimated directly from the number of atoms of the various elements which are collected. In some cases this information is all that is needed, as the ability to sample volumes 1–10 nm^3 permits the identification of phase chemistry on a very fine scale. Often, however, the variation of composition along the cylinder of analysis is investigated. In this case it is not only the number of atoms which is important, but also the order in which they are detected. The data chain representing the sequence of detected atoms can be investigated by subdivision into smaller blocks. The composition of each block can then be calculated. Changes in composition from block to block will then give direct information about the variation of composition with depth in the specimen (Fig. 17).

The volume represented by each block of data will depend on the way in which the subdivision has been performed. Typically, the compositions are calculated using a fixed number of atoms per block, say 50–100 atoms. While this method is the simplest, there is not necessarily a linear relationship between the position of a given block in the data sequence and the absolute depth in the specimen. The number of atoms detected per atomic layer varies from specimen to specimen, even for the same material, since the magnification varies with the end radius, and is also dependent on tip shape. There may also be significant variations from one phase to another within a given specimen of a multi-phase material, due to local magnification changes produced by the variation

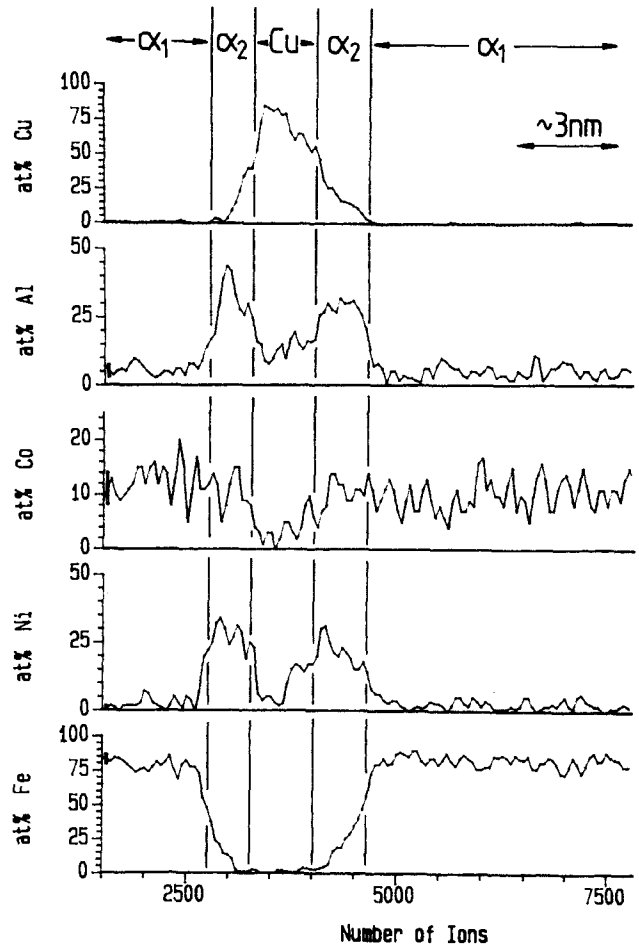


Figure 17. Atom probe composition–depth profile through the permanent magnet alloy Alnico 2 (Fe-20.8 at.%Al-14.8%Ni-11.3 Co-5.6%Cu). This material has a three-phase microstructure, consisting mostly of α_1 (Fe–Co rich) and α_2 (Al–Ni–Cu rich) phases, with a small volume fraction of copper precipitates formed within the α_2 phase. Reproduced from *Journal of Microscopy*, Volume 154, p. 215 (1989) by courtesy of the publishers.

in local radius discussed above. To obtain an absolute depth scale, the analysis is best performed with a small quantity of image gas present. The progress of field evaporation can then be monitored by following the collapse of rings on a known crystallographic pole. The size of each block in the composition profile can now be defined on the basis of the atoms detected during the removal of a fixed number of atomic planes (typically 2–5). Atomic layer depth resolution is obtained for composition variations perpendicular to the direction of analysis, but for changes of composition in directions away from the local surface normal the resolution will be reduced. For example, if an interface lies at an angle to the direction of analysis, the finite diameter of the analysis cylinder will lead to a depth profile which shows an apparently diffuse interface.

One important application of the single atom layer resolution of the atom probe is in studies of site occupancy in ordered materials. In this type of experiment, the analysis region is selected to lie just within the top layer of the superlattice pole in the FIM image. During the analysis, atoms are

detected from only one superlattice plane at a time, and the ordering can be studied directly. The conditions for such an experiment need to be selected carefully, if the measurement of ordering is to be accurate. Preferential field evaporation or retention of individual species must be minimized, and this requires the specimen temperature, pulse fraction, and field evaporation rate to be set more accurately than is usually necessary for routine analysis. Figure 18 shows an example from the analysis of Ni_3Al with a variety of additions, where the data are represented in the form of a 'ladder' diagram. This is a vector plot, in which the sequence of detection of the major species is displayed by a movement of the plot horizontally for one species, and vertically for the other. The composition of individual atomic planes in the material is given directly by the local slope of the resultant line. In Fig. 18, the ladder diagram is divided into regions of 45° slope with horizontal lines in between, corresponding to the analysis of mixed Ni + Al planes and pure Ni planes, respectively. Atoms of additional elements (Co, Fe, Hf) are represented by arrows, and show whether the solutes preferentially occupy sites in one plane or the other. Figure 18 indicates that both Fe and Hf are found predominantly in the mixed planes.

2.3.5 Analysis of Semiconductor Materials

For highly conductive materials such as metals, the use of voltage pulses to generate field evaporation is very effective, and most atom probe analysis is carried out in this way. However, in propagation through less conductive materials, such as semiconductors, the pulse can be severely attenuated. Even where sufficient amplitude does reach the apex of the specimen, the mechanical shock associated with the pulse can lead to brittle facture. In the analysis of semiconductors, therefore, voltage pulses are usually replaced by nanosecond laser pulses, focused on the apex of the specimen. This version of the atom probe is termed the pulsed laser atom probe (PLAP) [9]. The laser heating momentarily raises the temperature of the specimen from the base temperature to around 300 K, producing field evaporation under the influence of the applied d.c. voltage. This method has proved very useful in the analysis of a range of semiconducting materials and also for analysis of conducting ceramics. It should be emphasized that removal of material in the PLAP is by field evaporation, as in the atom probe, and not by thermal evaporation. Hence, the spatial resolution obtained in PLAP analysis is similar to that for the voltage pulsed atom probe. The technique should not be confused with laser ionization mass analysis (LIMA) in which material is removed from the specimen by thermal desorption at higher laser powers and much higher surface temperatures.

2.4 Three-Dimensional Atom Probes

The use of a probe aperture to define the lateral resolution in the atom probe severely restricts the total volume of

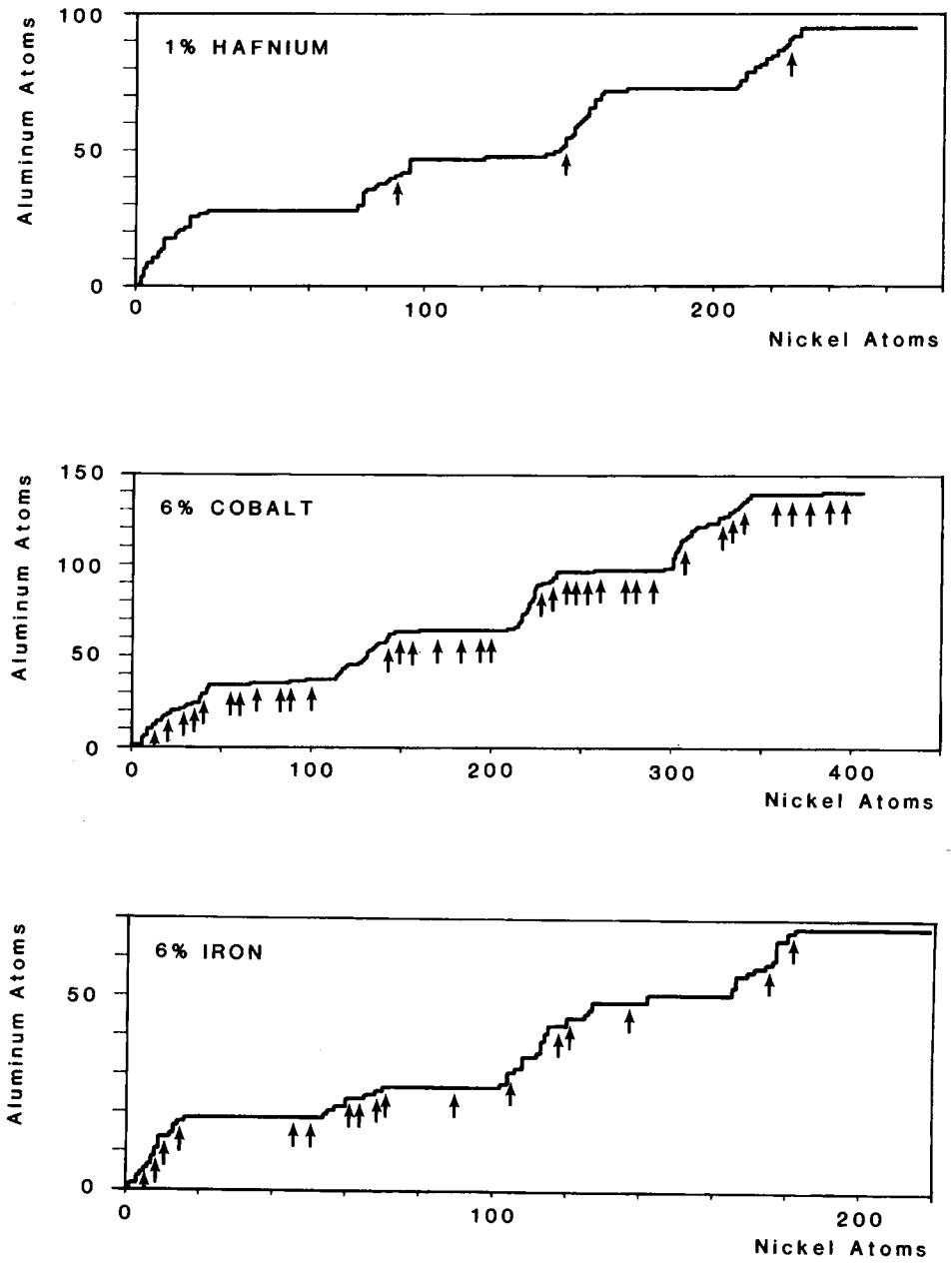


Figure 18. Ladder diagram showing the atom probe analysis along the (100) direction in $L1_2$-ordered nickel aluminum alloys (based on Ni_3Al) with additions of 1 at.%Hf, 6 at.%Co or 6 at.%Fe, showing the measurement of ordering along the superlattice direction, and the site occupation of the impurity atoms. The individual planes in the material are represented by regions of different slope on the plot, the horizontal lines being due to pure Ni planes, and a 45° slope representing a mixed Ni + Al plane. The position of the impurity atoms in the detection sequence is shown by arrows. Reproduced courtesy of Dr M. K. Miller.

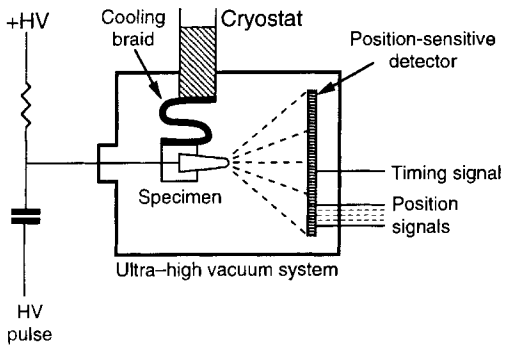


Figure 19. Schematic diagram of a three-dimensional atom probe. Flight times are measured over the short (100–600 mm) flight path between specimen and detector, and the position of impact of each ion is determined from the position-sensitive detector.

material which can be analyzed in any sample. This design also requires the accurate alignment of the probe aperture when a selected-area analysis is being performed. A recent development in the field of atom probe microanalysis has essentially removed these major limitations, by combining the time-of-flight mass spectrometry of the atom probe with position-sensing (Fig. 19). In this arrangement, the original position of the evaporated atom is inferred from the position of detection, in the same way that atom positions are determined from the image spots in a field ion micrograph. This eliminates the need for the selection aperture, and a larger area may be analyzed without sacrificing lateral resolution. Indeed, the spatial resolution obtained with position-sensitive detectors is better than that achieved with the probe hole aperture. The lateral resolution obtained with this instrument is typically limited to ± 0.5 nm by trajectory aberrations during field evaporation. The analysis proceeds by continued removal and detection of individual ions, atomic layer by atomic layer through the material. By reconstructing the position of the detected atoms within each layer, as well as the sequence of atomic layers, the original three-dimensional distribution of the elements can be displayed. This novel instrument design has therefore been termed the three-dimensional atom probe (3DAP).

2.4.1 Position-Sensing Schemes

There are a number of 3DAP designs, differing principally by the type of position-sensitive detector used. The first 3DAP was the position-sensitive atom probe (PoSAP), developed by Cerezo et al. [10]. In this instrument, the position-sensitive detector was based around a double MCP assembly with a wedge-and-strip anode. The wedge-and-strip anode was originally developed by Anger and coworkers [11] and works by dividing the charge incident from the MCPs between three electrodes in such a way as to encode the centroid of the charge cloud. This provides a simple method of position sensing, but one which is only effective for single ions. If more than one ion arrives within the time window for charge measurement (about 250 ns) then the position information is lost. This reduces the overall detection efficiency, and imposes analysis conditions which minimize the number of field evaporation pulses producing more than one detected ion, limiting the overall speed of analysis. Subsequently, Blavette et al. [12] developed their tomographic atom probe (TAP) which employs a multistrike detection system. In this detector, the electron cloud generated by ions striking the

MCPs is incident on a square anode array of 10×10 elements. A cloud will typically cover four elements in the array, and the ratio of the charge collected on each of the elements is used to calculate the centroid of the cloud. If two ions with sufficient spatial separation are incident on the detector, the position of each can be calculated in a simple manner. Kelly et al. [13] have proposed a similar multistrike detection system, using the principle of charge division in a conductive anode.

Another approach to position determination is to use a detector consisting of MCPs and a phosphor screen, and to couple the detector optically to an external position sensing detector. This makes the detector within the vacuum system much simpler, and the basic instrument configuration essentially that of an imaging atom probe, first developed by Panitz [14]. An additional advantage to this arrangement is that the detector can be used to obtain a direct field ion image of the analysis area. In the optical atom probe constructed by Miller [15, 16], the detector is fiber optic coupled to an image-intensified diode array camera. From the image acquired by the camera, individual ion positions are indicated by single spots of light. Again, this is essentially a single-event detection system, since if two different ions are detected on a single field evaporation pulse, it is not possible to ascertain which ion corresponds to which position. Cerezo et al. [17] have developed an optically-coupled multistrike detector system, called the optical PoSAP, which combines an image-intensified camera with an anode array photomultiplier tube. The image from the detector is split by a partially reflective mirror, and focused onto both the photomultiplier tube and the camera. A fast phosphor is used in the detector, so that the signals from the elements of the anode array can be used for flight-time measurements. Each element in the array corresponds to a certain area of the detector, so that a timing signal from one of the elements can easily be correlated with the exact impact position determined from the image acquired by the camera.

2.4.2 Mass Resolution in the Three-Dimensional Atom Probe

The first 3DAP designs used a straight flight path, and so suffered from the same limitations in mass resolution as the early probe-hole atom probe instruments. More recently, a 3DAP has been designed which uses a reflectron as an energy-compensating element [18]. The reflectron has the advantage that it acts as nothing more than a mirror in ion-optical terms, and so is free from image aberrations. However, care must be taken in the design of such a system, as the energy spreads are converted into position spreads, and therefore the initial energy spreads of ions must be kept to a minimum. The result is a 3DAP instrument with a mass resolution of $m/\Delta m = 600$.

2.4.3 Three-Dimensional Reconstruction of Atomic Chemistry

The combination of data provided by a 3DAP can obviously be used to construct both a mass analysis of the volume of material and also a map of the distribution

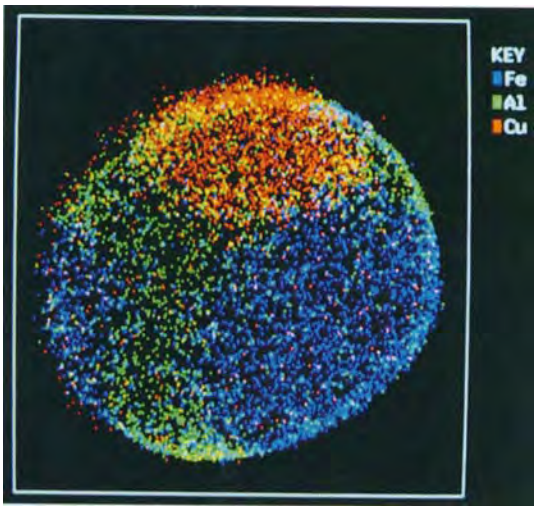


Figure 20. Element distributions from the PoSAP analysis of the permanent magnet alloy Alnico 2 (Fe-20.8 at.%Al-14.8%Ni-11.3 Co-5.6%Cu) from an area approximately 15 nm in diameter and 2–3 nm in depth. The three-phase nature of the material, as indicated in Fig. 17, is more clearly seen in these images. Note again how the copper precipitate is totally contained within the α_2 phase.

of an individual element on the specimen surface. It is also possible to map out any number of elements. An example of a PoSAP analysis of a complex alloy is shown in Fig. 20. The multielement mapping provided by the PoSAP gives a detailed insight into the atomic-scale phase chemistry of the material.

However, it is the ability to reconstruct the three-dimensional microstructure present in the specimen which is the most significant advantage of the 3DAP, and makes the technique almost unique. The TAP analysis from a nickel based superalloy is shown in Fig. 21 [19]. The γ and γ' phases are seen clearly in the atomic reconstructions, as is the segregation of Mo, B, and C to both interphase and grain boundaries. Note the individual (100) superlattice planes in the Al distributions, showing the ordering in the γ'-phase.

Miller et al. [20] and Hyde et al. [21, 22] have used the PoSAP to study phase separation in iron–chromium alloys, and compared the experimental results with computer simulations. With the PoSAP data, these authors have been able to compare directly atomic-scale data from experiment and the simulation, using chemical information, such as the amplitude of composition fluctuations, and also microstructural information.

2.5 Survey of Commercially Available Instrumentation

Field emission and field ion microscopes are relatively simple to construct, and many researchers build their own instruments from standard ultrahigh vacuum components. A commercial microscope for demonstrating field emission phenomena is available from Leybold Didactic GmbH, P.O. Box 1365, D-50330 Hürth, Germany. A compact table-top field ion microscope is manufactured by the Kitano Seiki Company Ltd, 7-17-3, Chuoh, Ohta-ku, Tokyo 143, Japan.

An energy-compensated atom probe instrument, the FIM100, which incorporates a Poschenrieder analyzer, is produced by VG Scientific, The Birches Industrial Estate, Imberhorne Lane, East Grinstead, West Sussex RH19 1UB, UK. Another energy-compensated atom probe, the APFIM 220, incorporating a reflectron analyzer, is available from Applied Microscopy, The Courtyard, Whitwick Business

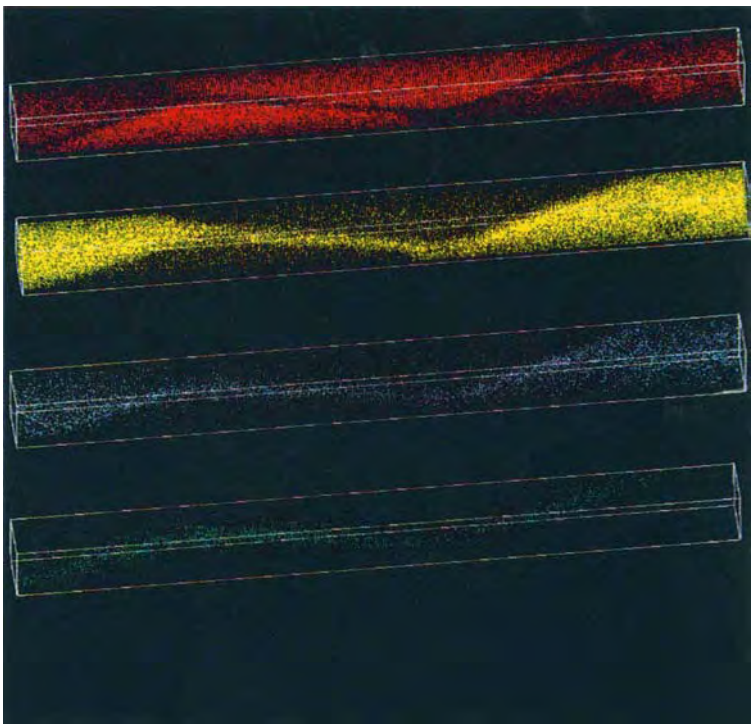


Figure 21. Tomographic atom probe analysis of the nickel-based superalloy Astroloy (Ni-8.5 at.%Al-16.0%Co-15.9%Cr-4.0%Ti-3.0%Mo-0.13%C-0.11%B-0.03%Zr) showing the atomic distributions of Al + Ti (red), Cr (yellow), Mo (blue) and B + C (green). The size of the volume analyzed is 12 nm × 12 nm × 98 nm. γ'-precipitates (Al rich) are clearly visible in the γ matrix (Cr rich), and Mo, B, and C are seen to segregate to grain and interphase boundaries. Reproduced courtesy of L. Letellier, M. Guttmannn and D. Blavette.

Park, Stenson Road, Whitwick, Leicestershire LE67 3JP, UK.

Position-sensitive detection systems for 3DAP are available from Kindbrisk Ltd, 8 Tilgarsley Road, Eynsham, Oxford OX8 1PP, UK (PoSAP and OPoSAP systems, with or without reflectron-based energy compensation) and from Cameca, 103 Boulevard Saint-Denis, B.P. 6, 92403 Courbevoie Cedex, France (TAP detector). Computer software for atom probe control, data acquisition and data analysis is also available from Kindbrisk Ltd.

2.6 References

[1] R. Gomer, *Field Emission and Field Ionization*, Harvard University Press, Harvard, CT **1961** (reprinted 1993, American Vacuum Society).

[2] E. W. Müller, T. T. Tsong, *Field Ion Microscopy, Principles and Applications*, Elsevier, Amsterdam **1969**.

[3] T. T. Tsong, *Atom Probe Field Ion Microscopy*, Cambridge University Press, Cambridge **1990**.

[4] M. K. Miller, A. Cerezo, M. G. Hetherington, G. D. W. Smith, *Atom Probe Field Ion Microscopy*, Oxford University Press, Oxford **1996**.

[5] T. T. Tsong, *Surf. Sci.* **1978,** *70*, 211.

[6] W. P. Poschenrieder, *Int. J. Mass Spectrom. Ion Phys.* **1972,** *9*, 357.

[7] V. I. Karataev, B. A. Mamyrin, D. V. Schmikk, *Sov. Phys.-Tech. Phys.* **1972,** *16*, 1177.

[8] M. Hellsing, L. Karlsson, H.-O. Andrén, H. Nordén, *J. Phys. E* **1985,** *18*, 920.

[9] G. L. Kellogg, T. T. Tsong, *J. Appl. Phys.* **1980,** *51*, 1184.

[10] A. Cerezo, T. J. Godfrey, G. D. W. Smith, *Rev. Sci. Instrum.* **1988,** *59*, 862.

[11] H. O. Anger, *U.S. Patent No. 3 209 201*, September 28, **1965**.

[12] D. Blavette, A. Bostel, J. M. Sarrau, B. Deconihout, A. Menand, *Rev. Sci. Instrum.* **1993,** *64*, 2911.

[13] T. F. Kelly, D. C. Mancini, J. J. McCarthy, N. A. Zreiba, *Surf. Sci.* **1991,** *246*, 396.

[14] J. A. Panitz, *J. Vac. Sci. Technol.* **1974,** *11*, 206.

[15] M. K. Miller, *Surf. Sci.* **1991,** *246*, 428.
[16] M. K. Miller, *Surf. Sci.* **1992,** *266*, 494.
[17] A. Cerezo, T. J. Godfrey, J. M. Hyde, S. J. Sijbrandij, G. D. W. Smith, *Appl. Surf. Sci.* **1994,** *76/77*, 374.
[18] S. J. Sijbrandij, A. Cerezo, T. J. Godfrey, G. D. W. Smith, *Appl. Surf. Sci.* **1996,** *94/95*, 428.
[19] L. Letellier, M. Guttmann, D. Blavette, *Phil. Mag. Lett.* **1994,** *70*, 189–194.
[20] M. K. Miller, J. M. Hyde, M. G. Hetherington, A. Cerezo, G. D. W. Smith, C. M. Elliott, *Acta Met. Mater.* **1995,** *43*, 3385.
[21] J. M. Hyde, M. K. Miller, M. G. Hetherington, A. Cerezo, G. D. W. Smith, C. M. Elliott, *Acta Met. Mater.* **1995,** *43*, 3403.
[22] J. M. Hyde, M. K. Miller, M. G. Hetherington, A. Cerezo, G. D. W. Smith, C. M. Elliott, *Acta Met. Mater.* **1995,** *43*, 3415.

Part VII

Scanning Point Probe Techniques

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

General Introduction

Scanning near-field probe microscopy (SNPM) has developed from scanning tunneling microscopy (STM), which was invented in 1981 by G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel at the IBM Zürich Research Laboratory [1–3]. SNPM combines three important concepts: scanning, point probing, and near-field operation [4]. Scanning is achieved by means of piezoelectric drives which allow the positioning and raster scanning of a point probe relative to a sample surface to be investigated with subatomic accuracy. Nonlinearities due to piezoelectric hysteresis and creep usually have to be corrected electronically or by software to prevent image distortions. Point probing allows local information to be obtained about the physical and chemical properties of a sample surface, which facilitates the investigation of site-specific sample properties. The point probe is brought in close proximity to the sample at a distance s which is smaller than some characteristic wavelength λ of a particular type of interaction between the probe tip and the sample. [In the case of STM, λ would be the electron wavelength whereas for scanning near-field optical microscopy (SNOM), λ would be the optical wavelength.] In this so-called near-field regime (where $s \leqslant \lambda$), the spatial resolution that can be achieved is no longer limited by diffraction, but rather by geometrical parameters: the distance s between the point probe and the sample surface, and the effective radius of curvature R of the point probe. SNPM is therefore an exceptional type of microscopy because it works without lenses (in contrast to optical and electron microscopy), and achieves 'super resolution' beyond the Abbé limit. Another important feature of SNPM is that it can be operated in air and in liquids as well as in vacuum, which offers novel opportunities for high-resolution studies of the structure and processes at solid/fluid interfaces. In particular, in situ electrochemical studies and in vivo investigations of biological specimens at unprecedented real-space resolutions have become some of the more recent intense fields of application, besides surface science studies under ultra-high-vacuum conditions.

1 Scanning Tunneling Microscopy

1.1 Introduction

Scanning tunneling microscopy (STM) was the first near-field microscopy technique to be developed. It is based on vacuum tunneling of electrons between an electrically conducting point probe and an electrically conducting sample (metal or doped semiconductor). The schematic set-up for an STM experiment is shown in Fig. 1. An atomically sharp probe tip is brought within a distance of only a few angstroms ($1\,\text{Å} = 0.1\,\text{nm} = 10^{-10}\,\text{m}$) from a sample surface by means of a piezoelectric drive in the z-direction (normal to the sample surface). If a bias voltage U has been applied between the tip and sample, a tunneling current I will flow due to the quantum mechanical tunneling effect even before mechanical point contact is reached. Since the tunneling current is strongly (exponentially) dependent on the tip–surface separation, it can be used very efficiently for distance control. By scanning the tip over the sample surface while keeping the tunneling current constant by means of an electronic feedback circuit, the surface contours can be traced by the tip. A quasi-three-dimensional 'topographic' image $z(x,y)$ of the sample surface is gained by monitoring the vertical position z of the tip as a function of the lateral position (x,y), which is controlled by piezoelectric drives in the x- and y-directions. The position (x,y,z) of the tip can be calculated based on the known sensitivities of the piezoelectric drives in the x-, y-, and z-directions and the corresponding driving voltages U_x, U_y, and U_z.

In addition to its use to control the tip–surface separation, the tunneling current contains valuable information about the local electronic properties of the sample surface and—to some extent—of the tip as well, which is unwanted in most cases. In the following, constant-current topographs and the various dependencies on experimental and sample-specific parameters will systematically be discussed. This will allow classification of the information which can be extracted from STM experiments.

1.2 Topographic Imaging in the Constant-Current Mode

Within Bardeen's transfer Hamiltonian formalism [5], the tunneling current I can be evaluated from the first-order time-dependent perturbation theory

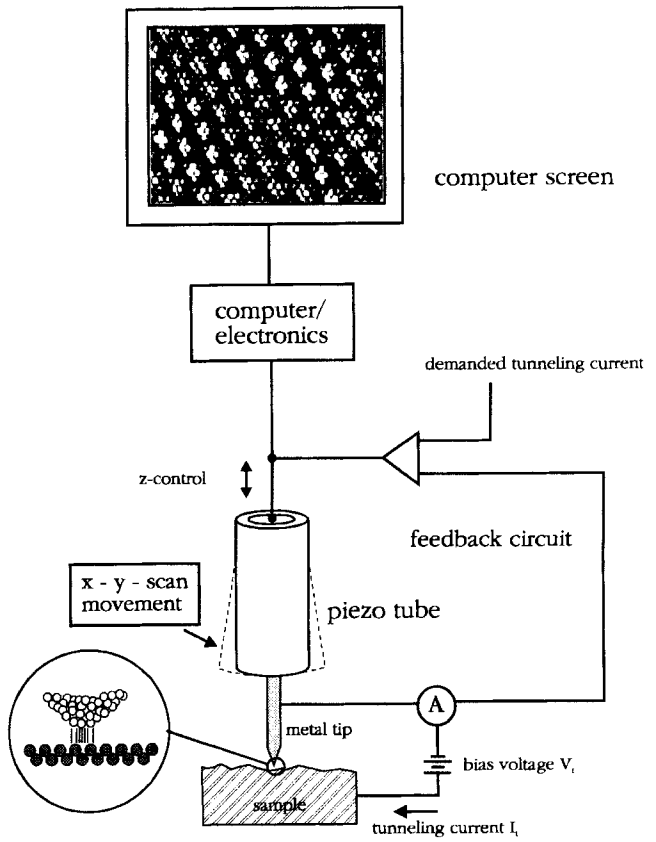


Figure 1. Schematic set-up for STM.

according to

$$I = \frac{2\pi e}{\hbar}\sum_{\mu,\nu}[f(E_\mu)(1-f(E_\nu+eU)) - f(E_\nu+eU)(1-f(E_\mu))] \times |M_{\mu,\nu}|^2\delta(E_\nu - E_\mu) \quad (1)$$

where $f(E)$ is the Fermi function, U is the applied sample bias voltage, $M_{\mu,\nu}$ is the tunneling matrix element between the unperturbed electronic states ψ_μ of the tip and ψ_ν of the sample surface, and $E_\mu(E_\nu)$ is the energy of the state $\psi_\mu(\psi_\nu)$ in the absence of tunneling. The delta function describes the conservation of energy for the case of elastic tunneling. (Inelastic tunneling will be considered later; see Sec. 1.6 of this Chapter.) The essential problem is the calculation of the tunneling matrix element which, according to Bardeen [5], is given by

$$M_{\mu,\nu} = \frac{\hbar^2}{2m}\int \mathrm{d}S\,(\psi_\mu^*\nabla\psi_\nu - \psi_\nu\nabla\psi_\mu^*) \quad (2)$$

where the integral has to be evaluated over any surface lying entirely within the vacuum barrier region separating the two electrodes. The quantity in parentheses can be identified as a current density $j_{\mu,\nu}$. To derive the matrix element $M_{\mu,\nu}$ from Eq. (2), explicit expressions for the wave functions ψ_μ and ψ_ν of the tip and sample

surface are required. Unfortunately, the atomic structure of the tip is generally not known. Therefore, a model tip wave function has to be assumed for calculation of the tunneling current.

Tersoff and Hamann [6, 7], who first applied the transfer Hamiltonian approach to STM, used the simplest possible model for the tip with a local spherical symmetry. In this model, the tunneling matrix element is evaluated for a s-type tip wave function, whereas contributions from tip wave functions with angular dependence (orbital quantum number $l \neq 0$) have been neglected. Tersoff and Hamann considered the limits of low temperature and small applied bias voltage, for which the tunneling current becomes

$$I = \frac{2\pi e}{\hbar} U \sum_{\mu,\nu} |M_{\mu,\nu}|^2 \times \delta(E_\nu - E_F)\delta(E_\mu - E_F) \quad (3)$$

where E_F is the Fermi energy. Within the s-wave approximation for the tip, the following expression for the tunneling current is finally obtained

$$I \propto U n_t(E_F) \exp(2\kappa R) \times \sum_\nu |\psi_\nu(\bar{r}_0)|^2 \delta(E_\nu - E_F) \quad (4)$$

with the decay rate $\kappa = (2m\phi)^{1/2}/\hbar$, where ϕ is the density of states at the Fermi level for the tip, R is the effective tip radius, and $\bar{r}_0$ is the center of curvature of the tip. The quantity

$$n_s(E_F, \bar{r}_0) = \sum_\nu |\psi_\nu(\bar{r}_0)|^2 \delta(E_\nu - E_F) \quad (5)$$

can be identified with the surface local density of states (LDOS) at the Fermi level E_F, that is, the charge density from electronic states at E_F, evaluated at the center of curvature r of the effective tip. The STM images obtained at low bias in the constant-current mode therefore represent contour maps of constant surface LDOS at E_F evaluated at the center of curvature of the effective tip, provided that the s-wave approximation for the tip can be justified. Since the wave functions decay exponentially in the z-direction normal to the surface toward the vacuum region,

$$\psi_\nu(\bar{r}) \propto \exp(-\kappa z) \quad (6)$$

it follows that

$$|\psi_\nu(\bar{r}_0)|^2 \propto \exp[-2\kappa(s + R)] \quad (7)$$

where s denotes the distance between the sample surface and the front end of the tip. Therefore, the tunneling current, given by Eq. (4), becomes exponentially dependent on the distance s, as mentioned in the introduction:

$$I \propto \exp(-2\kappa s) \quad (8)$$

The strong exponential distance dependence typically leads to an order-of-magnitude increase in the tunneling current for each angstrom decrease in the tip–surface separation.

Unfortunately, the simple interpretation of constant current STM images as given by Tersoff and Hamann is not valid for high bias or for tip wave functions with angular dependence.

1.2.1 Effects of Finite Bias

The applied bias voltage enters through the summation of states which can contribute to the tunneling current. Additionally, a finite bias can lead to a distortion of the tip and sample surface wave functions

ψ_μ and ψ_ν as well as to a modification of the energy eigenvalues E_μ and E_ν [8]. The derivation of these distorted tip and sample surface wave functions and energy eigenvalues under the presence of an applied bias is, however, a difficult problem. Therefore, as a first approximation, the undistorted zero-voltage wave functions and energy eigenvalues are usually taken. Consequently, the effect of a finite bias U only enters through a shift in energy of the undistorted surface wave functions or density of states relative to the tip by an amount eU. Under this approximation, the following expression for the tunneling current as a generalization of the result of Tersoff and Hamann may be used:

$$I \propto \int_0^{eU} n_t(\pm eU \pm \mathscr{E}) n_s(\mathscr{E}, \bar{r}_0)\, d\mathscr{E} \qquad (9)$$

where $n_t(\mathscr{E})$ is the density of states for the tip and $n_s(\mathscr{E}, \bar{r}_0)$ is the density of states for the sample surface evaluated at the center of curvature $\bar{r}_0$ of the effective tip. All energies $\mathscr{E}$ are measured with respect to the Fermi level. One can now make the following approximation motivated by a generalization of Eq. (5) together with Eq. (7):

$$n_s(\mathscr{E}, \bar{r}_0) \propto n_s(\mathscr{E}) \exp\left\{-2(s+R) \times \left[\frac{2m}{\hbar^2}\left(\frac{\phi_t + \phi_s}{2} + \frac{eU}{2} - \mathscr{E}\right)\right]^{1/2}\right\} \qquad (10)$$

where a WKB-type expression for the decay rate κ in the exponential term has been used. $\phi_t(\phi_s)$ denotes the tip (sample surface) work function. Finally, one obtains

$$I \propto \int_0^{eU} n_t(\pm eU \pm \mathscr{E}) n_s(\mathscr{E}) T(\mathscr{E}, eU)\, d\mathscr{E} \qquad (11)$$

with an energy- and bias-dependent transmission coefficient $T(\mathscr{E}, eU)$ given by

$$T(\mathscr{E}, eU) = \exp\left\{-2(s+R) \times \left[\frac{2m}{\hbar^2}\left(\frac{\phi_t + \phi_s}{2} + \frac{eU}{2} + \mathscr{E}\right)\right]^{1/2}\right\} \qquad (12)$$

In Eqs. (11) and (12), matrix element effects in tunneling are expressed in terms of a modified decay rate κ including a dependence on energy E and applied bias voltage U. The expression (12) for the transmission coefficient neglects image potential effects as well as the dependence of the transmission probability on parallel momentum. This can be included by an increasingly more accurate approximation for the decay rate κ.

1.2.2 Effects of Tip Wave Functions with Angular Dependence

STM tips are usually made from tungsten or platinum–iridium alloy wire. For these materials, the density of states at the Fermi level is dominated by d-states rather than by s-states. Indeed, first-principle calculations of the electronic states of several types of tungsten clusters used to model the STM tip revealed the existence of dangling-bond states near the Fermi level at the apex atom which can be ascribed to d_{z^2} states [9]. Evaluation of the tunneling current according to Eqs. (1) and (2) now requires calculation of the tunneling matrix element for tip wave functions with angular dependence ($l \neq 0$). Chen [10] has shown that generally the tunneling matrix element can simply be obtained from a 'derivative rule'. The angle dependence of the tip wave function in terms of

x, y, and z has to be replaced according to

$$x \rightarrow \frac{\partial}{\partial x}$$
$$y \rightarrow \frac{\partial}{\partial y} \qquad (13)$$
$$z \rightarrow \frac{\partial}{\partial z}$$

where the derivatives have to act on the sample surface wave function at the center of the apex atom. For instance, the tunneling matrix element for a p_z tip state is proportional to the z-derivative of the sample surface wave function at the center of the apex atom at r_0.

In terms of a microscopic view of the STM imaging mechanism [11] illustrated in Fig. 2, a dangling-bond state at the tip apex atom is scanned over a two-dimensional array of atomic-like states at the sample surface. Overlap of the tip state with the atomic-like states on the sample surface generates a tunneling conductance which depends on the relative position of the tip state and the sample state. The atomic corrugation Δz depends on the spatial distribution as well as on the type of tip and sample surface states. Generally, for non-s-wave tip states, the tip apex atom follows a contour, determined by the derivatives of the sample surface wave functions, which exhibit much stronger atomic corrugation than the contour of constant surface LDOS at E_F.

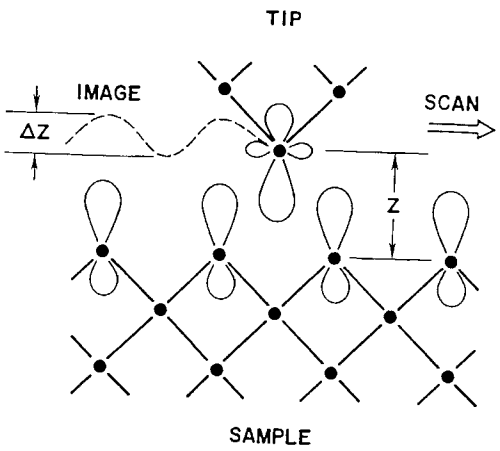


Figure 2. Microscopic view of the STM imaging mechanism. (Image courtesy of Chen [11].)

1.2.3 Imaging of Adsorbates

The transfer Hamiltonian approach as used by Tersoff and Hamann has further limitations. First, it is a perturbative treatment of tunneling, being appropriate only for weakly overlapping electronic states of the two electrodes. Secondly, this approach suffers from the fact that assumptions for the tip and sample surface wave functions have to be made in order to derive the tunneling current.

As an alternative, Lang [12, 13] has calculated the tunneling current between two planar metal electrodes with adsorbed atoms where the wave functions for the electrodes have been obtained self-consistently within density functional theory. In Fig. 3a the calculated current density distribution from a single sodium atom adsorbed at its equilibrium distance on one of the two metal electrodes is shown. The plot illustrates how spatially localized the tunneling current is. By scanning one adsorbed atom (taken as the tip) past another adsorbed atom (taken as the sample), the vertical tip displacement versus the lateral position can be evaluated under the constant-current condition [13]. In Fig. 3b, constant-current scans at low bias of a sodium tip atom past three different sample adatoms (sodium, sulfur,

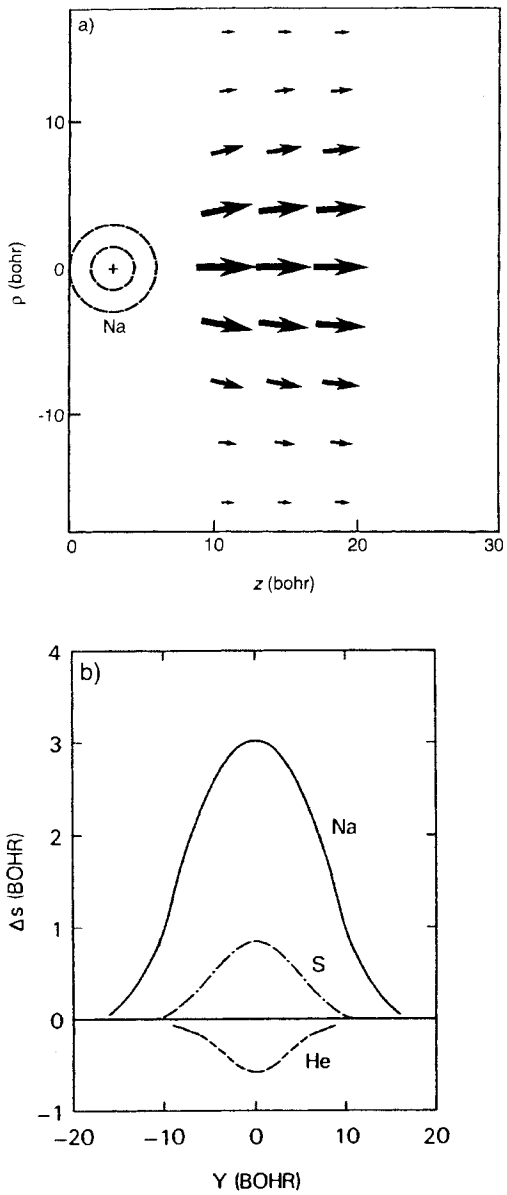


Figure 3. (a) Current density for the case in which a sodium atom is adsorbed on the left electrode. The length and thickness of the arrows are proportional to $\ln(ej/j_0)$ evaluated at the spatial position corresponding to the center of the arrow (1 bohr = 0.529 Å). (Image courtesy of Lang [14].) (b) Change in tip distance versus lateral separation for constant tunnel current. The tip atom is sodium; sample adatoms are sodium, sulfur, and helium. (Image courtesy of Lang [13].)

and helium) are shown. Most striking is the negative tip displacement for adsorbed helium. The closed valence shell of helium is very much lower in energy with respect to the Fermi level, and its only effect is to polarize metal states away from E_F, thereby producing a decrease in the Fermi level state density. This results in a reduced tunneling current flow, that is, a negative tip displacement in a constant-current scan. This example illustrates nicely that 'bumps' or 'holes' in 'topographic' STM images may not correspond to the presence or absence of surface atoms, respectively—sometimes even the reverse is true.

1.2.4 Spatial Resolution in Constant-Current Topographs

According to Tersoff and Hamann [7], an STM corrugation amplitude, or corrugation in brief, Δ, may be defined by

$$\Delta := z_+ - z_- \tag{14}$$

where z_+ and z_- denote the extremal values of the z-displacement of the tip in a constant-current scan. This corrugation Δ decreases exponentially with distance z from the surface,

$$\Delta \propto \exp(-\gamma z) \tag{15}$$

where the decay rate γ is very sensitive to the surface lattice constant because it depends quadratically on the corresponding Fourier component G in the expansion of the surface charge density,

$$\gamma \propto \tfrac{1}{4}\kappa^{-1}G^2 \tag{16}$$

with $\kappa^2 = 2m\phi/\hbar^2$. Consequently, only the lowest non-zero Fourier component determines the corrugation at sufficiently large

distances. Tersoff and Hamann [6, 7] argued that suppression of higher Fourier components in their expression for the tunneling current between a spherical tip of radius R and a sample surface at a distance s from the front end of the tip is equivalent to a spatial resolution determined by

$$\left(\frac{R+s}{\kappa}\right)^{1/2} \tag{17}$$

According to this expression, the lateral resolution in STM is determined by the geometrical parameters R and s, rather than by the wavelength of the tunneling electrons. This is characteristic for near-field microscopes which are operated at distances between the probe tip and the sample surface that are small compared with the wavelength, as mentioned in the introduction. For STM, typical tip–surface separations are 3–10 Å, whereas the wavelength of tunneling electrons typically varies in the range 12–120 Å for an applied bias voltage of 0.01–1 V.

The expression, Eq. (17), for the lateral resolution in constant-current STM images implies that high spatial resolution is obtained with a small radius of curvature of the effective tip and at a small tip–surface separation, that is, at low tunneling gap resistance. Both dependencies have been verified experimentally. The dependence of the measured corrugation on the radius of curvature of the effective tip was studied by combined STM–FIM (field ion microscopy) experiments [15] where the obtained STM results could be correlated directly with the size of the effective tip as revealed by FIM (Fig. 4). As a direct consequence, measured absolute values for the corrugation Δ are meaningless if the microscopic structure of the tip is not known.

The dependence of the measured corrugation on the tip–surface separation has experimentally been studied by systematic variation of the tunneling gap resistance [16, 17]. In particular, the suppression of higher Fourier components in the

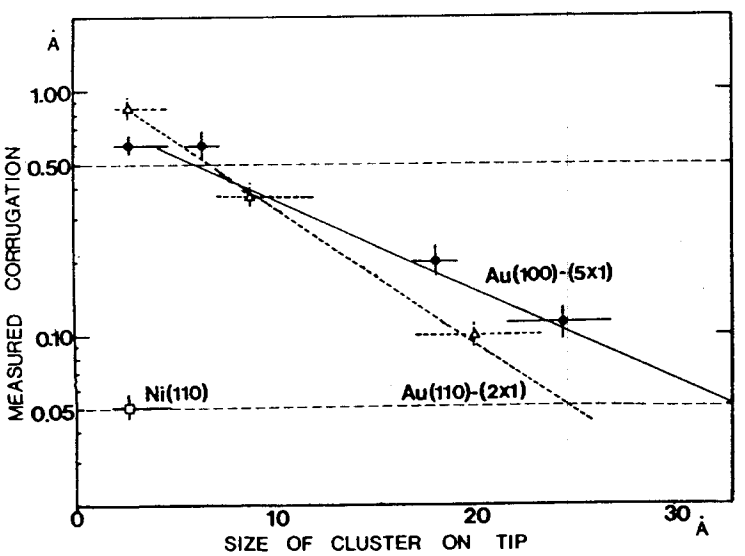


Figure 4. Dependence of the measured corrugation on the size of the cluster on the tip for Au(110) (2 × 1) and Au(100) (5 × 1) reconstructed surfaces. (Image courtesy of Kuk [15].)

expansion of the surface charge density with increasing tip–surface separation, as theoretically predicted by Tersoff and Hamann [6, 7], has experimentally been verified [17]. Figure 5 shows the influence of the chosen tip–surface separation on the spatial resolution achieved on a W(110)/C-(15 × 3) reconstructed surface. It is immediately apparent that the STM results can critically depend on the tip–surface separation, that is, on the tunneling gap resistance, particularly for surface structures with complex unit cell structure.

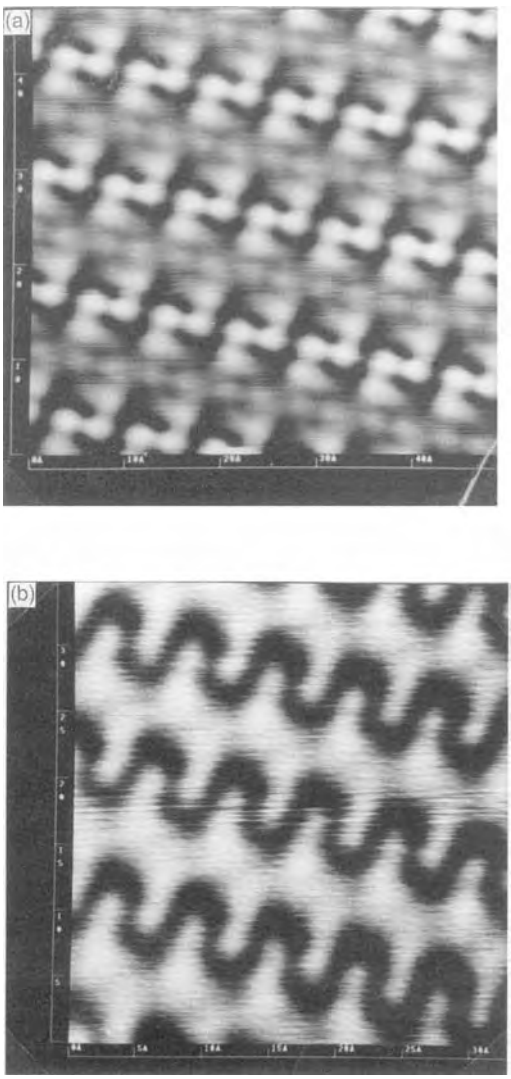


Figure 5. (a) STM topograph of the W(110)/C-R(15 × 3) reconstructed surface obtained with a tunneling gap resistance $R = 2.11 \times 10^6\ \Omega$. (b) Corresponding STM topograph with a tunneling gap resistance $R = 1.72 \times 10^{10}\ \Omega$ [17].

For close-packed metal surfaces, such as Au(111) [18] or Al(111) [16], atomic resolution could not be explained within the spherical tip model employed by Tersoff and Hamann. Baratoff [19] early on pointed out that the spatial resolution might be considerably improved compared with expression (17) if tunneling occurs via localized surface states or dangling bonds. Later, Chen [11, 20] systematically investigated the influence of different tip orbitals on the spatial resolution within a microscopic view of STM. The calculated enhancement of the tunneling matrix element by tip states with $l \neq 0$ was shown to lead to increased sensitivity to atomic-sized features with large wavevectors. For instance, a p_z tip state acts as a quadratic high-pass filter, whereas a d_{z^2} tip state acts as a quartic high-pass filter. Consequently, the resolution of STM can be considerably higher than predicted within the s-wave tip model. The spontaneous switching of the resolution often observed in or between atomic-resolution STM images can be explained by the fact that a very subtle change of the tip involving a change of the effective orbital can induce a tremendous difference in STM resolution. In conclusion, it is the orbital at the front end of the tip which mainly determines the spatial resolution in STM. A p_z orbital typical for elemental semiconductors or a d_{z^2} orbital from d-band metals are

most favorable. Therefore, 'tip-sharpening procedures' have to aim at bringing such favorable orbitals to the front of the tip [11].

1.3 Local Tunneling Barrier Height

According to Eq. (8), the tunneling current I was found to depend exponentially on the tip–surface separation s:

$$I \propto \exp(-2\kappa s)$$

with a decay rate κ given by

$$\kappa = \frac{(2m\phi)^{1/2}}{\hbar}$$

where ϕ is an effective local potential barrier height. So far, ϕ was assumed to be laterally uniform. In reality, ϕ exhibits spatial variations which can yield additional information about the sample surface under investigation.

Motivated by Eq. (8), an apparent local barrier height is usually defined by

$$\phi_A = \frac{\hbar^2}{8m}\left(\frac{\mathrm{d}\ln I}{\mathrm{d}s}\right)^2 \tag{18}$$

For large tip–surface separations outside the effective range of image forces, it is clear that ϕ has to approach the surface local work function ϕ_s, which is defined as the work needed to remove an electron from the Fermi level of the solid to a position somewhat outside of the surface where image force effects can be neglected. However, for small tip–surface separations (5–10 Å), image potential effects certainly have to be considered. By assuming a model potential [21]

$$\phi(d) = \phi_0 - \frac{\alpha}{d} \tag{19}$$

where ϕ_0 is the average work function of the sample surface and the probe tip $[\phi_0 = (\phi_s + \phi_t)/2]$, and d is the distance between the two image planes ($d \approx s - 1.5$ Å), the distance dependence of the tunneling current can be calculated:

$$\frac{\mathrm{d}\ln I}{\mathrm{d}s} = -\frac{2(2m)^{1/2}}{\hbar} \times \phi_0^{1/2}\left[1 + \frac{\alpha^2}{8\phi_0^2 d^2} + \mathcal{O}\left(\frac{1}{d^3}\right)\right] \tag{20}$$

As can be seen from Eq. (20), the first-order term in $1/d$, although present in the potential $\phi(d)$, cancels exactly in the expression for $\mathrm{d}\ln I/\mathrm{d}s$. The second-order term in $1/d$ usually contributes only a few percent of the zero-order term and can therefore be neglected to a first approximation. As a consequence, one finds

$$\frac{\mathrm{d}\ln I}{\mathrm{d}s} \approx \text{const.} \tag{21}$$

and

$$\phi_A \approx \phi_0 = \text{const.}$$

This means that the presence of the image potential does not show up in the distance dependence of the tunneling current although the absolute values of the current are drastically affected by the presence of the image potential. The distance independence of the apparent barrier height deduced from the $\ln I$–s relation (Fig. 6) has been verified experimentally as well as by more detailed theoretical analysis [22].

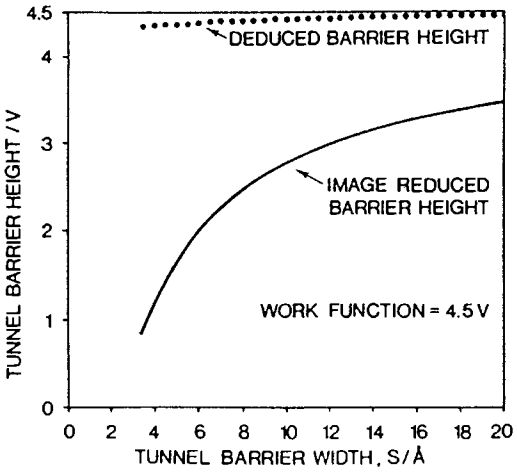


Figure 6. The image-reduced mean barrier height (full line) and the apparent barrier height deduced from the $\ln I$–s relation for this barrier (dotted line). The work function used in the calculation is 4.5 eV. It can be seen that the apparent barrier height is always within 0.2 eV of the work function despite the collapse of the image-reduced barrier. (Image courtesy of Coombs [22].)

1.3.1 Local Tunneling Barrier Height Measurements at Fixed Surface Locations

According to Eq. (18), the apparent barrier height ϕ_A can be determined locally by measuring the slope of $\ln I$–s characteristics at a fixed sample bias voltage U and at a fixed sample surface location. To demonstrate vacuum tunneling it is necessary to obtain reasonably high values for ϕ_A of several electronvolts in addition to verifying the exponential dependence of the current on the tip–surface separation.

Alternatively, the apparent barrier height can be deduced from the slope of local $\ln U$–s characteristics in a low applied bias voltage range and at a fixed tunneling current. In the low-bias regime, the tunnel junction exhibits Ohmic behavior, as found earlier (4):

$$I \propto U \exp(-2\kappa s) \tag{22}$$

Therefore, one obtains

$$\phi_A = \frac{\hbar^2}{8m}\left(\frac{\mathrm{d} \ln U}{\mathrm{d}s}\right)^2 \tag{23}$$

at constant current.

1.3.2 Spatially Resolved Local Tunneling Barrier Height Measurements

The experimental determination of the spatially resolved local tunneling barrier height $\phi_A(x, y)$ can be performed by modulating the tip–surface separation s by Δs while scanning at a constant average current I, with a modulation frequency ν_0 higher than the cut-off frequency of the feedback loop [23]. The modulation of $\ln I$ at ν_0 can be measured by a lock-in amplifier simultaneously with the corresponding constant-current topograph, and directly yields a signal proportional to the square root of the apparent barrier height via the relation

$$\frac{\Delta \ln I}{\Delta s} = -\frac{2\sqrt{2m}}{\hbar}\sqrt{\phi_A} \tag{24}$$

The apparent barrier height obtained in this way is not measured at a constant tip–surface separation s. Scanning at a constant average current (and at a constant applied bias voltage) implies that the product $\sqrt{\phi_A}s$ is kept constant, rather than s. However, since the spatial variation of ϕ_A is usually small (about 10% or less of the absolute value of

ϕ_A), and ϕ_A enters only under the square root, the spatial variation of $\phi_A(x,y)$ is usually measured almost at a constant tip–surface separation s.

Spatially resolved measurements of the apparent potential barrier height can yield information about spatial inhomogeneities of the local sample work function ϕ_s, which can be split into two contributions. The chemical component of ϕ_s is determined by the chemical nature and structure of the solid only, whereas the electrical component of ϕ_s depends on the chemical nature of the solid as well as on the surface crystallographic orientation. Therefore, spatially resolved measurements of ϕ_A can be used, for instance, to map chemical inhomogeneities at surfaces as well as different crystallographic facets of small crystallites.

On the atomic scale, it is more appropriate to relate the measured apparent barrier height with the decay rates of the wave functions describing the sample surface and the tip. Lateral variations of $\phi_A(x,y)$ then have to be interpreted as lateral variations in the decay rate of the surface wave function. As we know from Eq. (15), the measured surface atomic corrugation Δ in constant-current STM images is smoothed out exponentially with an increasing tip–surface separation s. This can only occur if the decay rate κ_p above a local protrusion in the topography is larger than the decay rate κ_d above a local depression. Consequently, the apparent barrier height above a local topographic protrusion has to be larger than the barrier height above a local depression. Therefore, atomically resolved apparent barrier height images closely reflect corresponding topographic constant-current images.

1.4 Tunneling Spectroscopy

Besides the distance and apparent barrier height dependence of the tunneling current there also exists a bias dependence which can be studied by various tunneling spectroscopic methods. For tunneling between metal electrodes in the low-bias limit, the tunneling current is found to be linearly proportional to the applied bias voltage [Eq. (4)]. For higher bias and particularly for semiconductor samples, the bias dependence of the tunneling current generally does not exhibit Ohmic behavior, and the constant-current STM images can depend critically on the applied bias (Fig. 7). Studying this bias dependence in detail allows extraction of various spectroscopic information at high spatial resolution, ultimately down to the atomic level. The spectroscopic capability of STM combined with its high spatial resolution is perhaps the most important feature of STM, and has been applied widely, particularly for investigation of semiconductor surfaces [24–26]. Figure 8 shows a simplified one-dimensional potential energy diagram at zero temperature for the system consisting of the tip (left electrode) and the sample (right electrode), which are separated by a small vacuum gap. For zero applied bias (Fig. 8b) the Fermi levels of tip and sample are equal at equilibrium. When a bias voltage U is applied to the sample, the main consequence is a rigid shift of the energy levels downward or upward in energy by an amount $|eU|$, depending on whether the polarity is positive (Fig. 8c) or negative (Fig. 8d). (As discussed in Sec. 1.2.1 of this Chapter, we neglect the distortions of the wave functions and the energy eigenvalues due to the

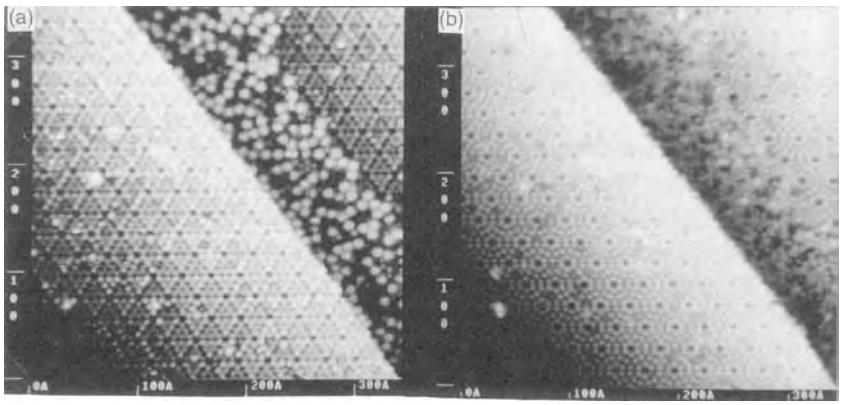


Figure 7. (a) STM topograph of the Si(111)7 × 7 reconstructed surface with a step along which molecules were found to be preferentially adsorbed. The image was taken with negative sample bias voltage polarity. (b) Corresponding STM image obtained with positive sample bias voltage polarity. The adsorbed molecules have become almost invisible. The Si(111)7 × 7 surface also appears different depending on the bias voltage polarity.

finite bias to a first approximation.) For positive sample bias, the net tunneling current arises from electrons that tunnel from the occupied states of the tip into unoccupied states of the sample (Fig. 8c), whereas at negative sample bias, electrons tunnel from occupied states of the sample into unoccupied states of the tip. Consequently, the bias polarity determines whether unoccupied or occupied sample electronic states are probed. It also becomes clear that the electronic structure of the tip enters as well, as is also obvious from Eq. (11) for the tunneling current.

By varying the amount of the applied bias voltage, one can select the electronic states that contribute to the tunneling current and, in principle, measure the local electronic density of states. For instance, the current increases strongly if the applied bias voltage allows the onset of tunneling into a maximum of the unoccupied sample electronic density of states. Therefore, the first derivative $\mathrm{d}I/\mathrm{d}U(U)$ reflects the electronic density of states to a first approximation. However, the energy and bias dependence of the transmission coefficient $T(\mathscr{E}, eU)$ has also to be considered. Since electrons in states with the highest energy 'see' the smallest effective barrier height, most of the tunneling current arises from electrons near the Fermi level of the negatively biased electrode. This has been indicated in Fig. 8 by arrows of differing size. The maximum in the transmission coefficient $T(\mathscr{E}, eU)$ given in Eq. (12) can be written as [26]

$$T_{\max}(U) = \exp\left\{-2(s+R) \times \left[\frac{2m}{\hbar^2}\left(\frac{\phi_t + \phi_s}{2} + \frac{|eU|}{2}\right)\right]^{1/2}\right\} \qquad (25)$$

The bias dependence of the transmission coefficient typically leads to an order-of-magnitude increase in the tunneling current for each volt increase in magnitude of the applied bias voltage. Since the transmission coefficient increases monotonically with the applied bias voltage, it contributes only a smoothly varying

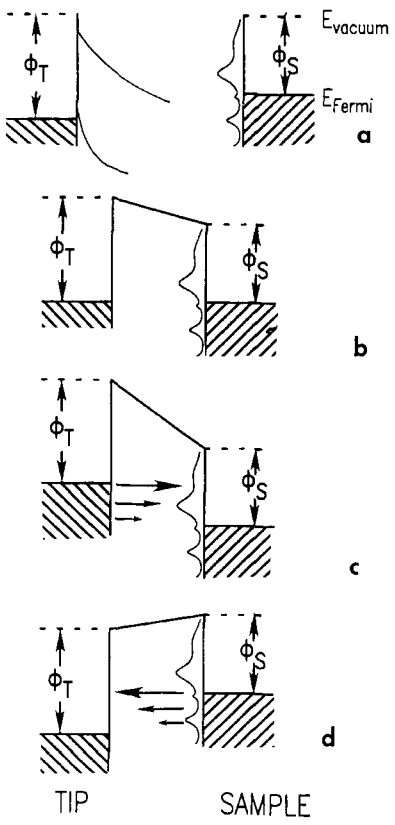


Figure 8. Energy level diagrams for the sample and the tip. (a) Independent sample and tip. (b) Sample and tip at equilibrium, separated by a small vacuum gap. (c) Positive sample bias: electrons tunnel from the tip to the sample. (d) Negative sample bias: electrons tunnel from the sample into the tip. (Image courtesy of Hamers [24].)

background on which the density-of-states information is superimposed.

As an important consequence of the dominant contribution of tunneling from states near the Fermi level of the negatively biased electrode, tunneling from the tip to the sample (Fig. 8c) mainly probes the empty states of the sample with negligible influence of the occupied states of the tip. On the other hand, tunneling from the sample to the tip is much more sensitive to the electronic structure of the empty states of the tip, which often prevents detailed spectroscopic STM studies of the occupied states of the sample [27].

1.4.1 Scanning Tunneling Spectroscopy at Constant Current

To perform scanning tunneling spectroscopy measurements, a high-frequency sinusoidal modulation voltage can be superimposed on the constant bias voltage applied between the tip and the sample. The modulation frequency is chosen higher than the cut-off frequency of the feedback loop, which keeps the average tunneling current constant. By recording the tunneling current modulation, which is in-phase with the applied bias voltage modulation, by means of a lock-in amplifier, a spatially resolved spectroscopy signal $\mathrm{d}I/\mathrm{d}U$ can be obtained simultaneously with the constant current image [28, 29]. Based on expression (11) for the tunneling current and by assuming $\mathrm{d}n_t/\mathrm{d}U \approx 0$, one obtains [24]

$$\frac{\mathrm{d}I}{\mathrm{d}U}(U) \propto e n_t(0) n_s(eU) T(eU, eU) + \int_0^{eU} n_t(\pm eU \mp \mathscr{E}) n_s(\mathscr{E}) \times \frac{\mathrm{d}T(\mathscr{E}, eU)}{\mathrm{d}U}\,\mathrm{d}\mathscr{E} \qquad (26)$$

At a fixed location, the increase of the transmission coefficient with applied bias voltage is smooth and monotonic. Therefore, structure in $\mathrm{d}I/\mathrm{d}U$ as a function of U can usually be attributed to structure in the state density via the first term in Eq. (26). However, interpretation of the spectroscopic data $\mathrm{d}I/\mathrm{d}U$ as a function of position

(x, y) is more complicated. As discussed in Sec. 1.3.2 of this Chapter, the apparent barrier height above a local topographic protrusion is larger, that is, the transmission coefficient is smaller, than above a local topographic depression. This spatial variation in the transmission coefficient shows up in spatially resolved measurements of $\mathrm{d}I/\mathrm{d}U$ as a 'background' that is essentially an 'inverted' constant-current topography. Therefore, spectroscopic images corresponding to the spatial variation of $\mathrm{d}I/\mathrm{d}U$ obtained in the constant current mode in fact contain a superposition of topographic and electronic structure information.

1.4.2 Local Spectroscopic Measurements at Constant Separation

To eliminate the influence of the z-dependence of the transmission coefficient, local $I-U$ characteristics can be measured at a fixed tip–sample separation. This is achieved by breaking the feedback circuit for a certain time interval at selected surface locations by means of a sample-and-hold amplifier, while local $I-U$ curves are recorded [30, 31]. The $I-U$ characteristics are usually repeated several times at each surface location and finally signal averaged. Since the feedback loop is inactive while sweeping the applied bias voltage, the tunneling current is allowed to become extremely small. Therefore, band gap states in semiconductors, for instance, can be probed without difficulties. The first derivative $\mathrm{d}I/\mathrm{d}U$ can be obtained from the measured $I-U$ curves by numerical differentiation. The dependence of the measured spectroscopic data on the value of the tunneling conductance I/U can be compensated by normalizing the differential conductance $\mathrm{d}I/\mathrm{d}U$ to the total conductance I/U. The normalized quantity $(\mathrm{d}I/\mathrm{d}U)/(I/U)$ reflects the electronic density of states reasonably well by minimizing the influence of the tip–sample separation [32]. However, the close resemblance of the $(\mathrm{d}\ln I/\mathrm{d}\ln U)-U$ curve to the electronic density of states is generally limited to the position of peaks while peak intensities can differ significantly.

1.4.3 Current Imaging Tunneling Spectroscopy

The measurements of local $I-U$ curves at a constant tip–sample separation can be extended to every pixel in an image, which allows performance of atomically resolved spectroscopic studies [33]. The method, denoted current imaging tunneling spectroscopy (CITS), also uses a sample-and-hold amplifier to alternately gate the feedback control system on and off. During the time of active feedback, a constant stabilization voltage U_0 is applied to the sample, and the tip height is adjusted to maintain a constant tunneling current. When the feedback system is deactivated, the applied sample bias voltage is linearly ramped between two preselected *values*, and the $I-U$ curve is measured at a fixed tip height. Afterwards, the applied bias voltage is set back to the chosen stabilization voltage U_0 and the feedback system is reactivated. By acquiring the $I-U$ curves rapidly while scanning the tip position at low speed, a constant-current topograph and spatially resolved $I-U$ characteristics

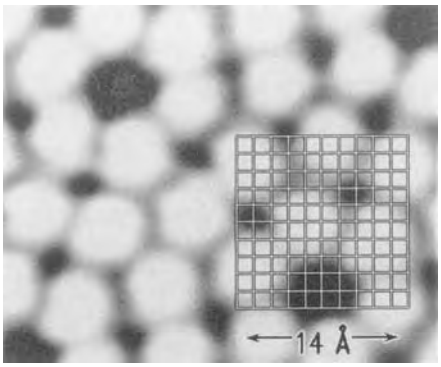


Figure 9. STM topograph of the unoccupied states of an Si(111)7 × 7 surface (sample bias 2 V). The atoms imaged are the top-layer Si adatoms (top). The grid encompasses a 14 Å × 14 Å area of this surface for which tunneling spectra have been obtained. The 100 tunneling spectra are plotted in the dI/dU form. Such spectral maps allow one not only to obtain the energies of the occupied (negative bias) and unoccupied (positive bias) states of particular atomic sites, but also to obtain information on the spatial extent of their wave functions (bottom). (Image courtesy of Avouris [34].)

can simultaneously be obtained. To increase the possible scan speed and to decrease the amount of data to be stored, one can predefine a coarse grid of pixels in the image at which local $I-U$ curves will be measured (Fig. 9). The ability to probe the local electronic structure down to atomic scale has great potential, for instance, for investigation of surface chemical reactivity on an atom-by-atom basis [34].

1.5 Spin-Polarized Scanning Tunneling Microscopy

Thus far, the dependence of the tunneling current on the tip–sample separation s, the local barrier height, and the applied sample bias voltage U has been considered:

$$I = I(s, \phi, U) \tag{27}$$

Accordingly, the corresponding modes of STM operation have been discussed: 'topographic' imaging, local barrier height imaging, and tunneling spectroscopy. However, the spin of the tunneling electrons and the additional spin dependence of the tunneling current, if magnetic electrodes are involved, have not yet been considered:

$$I = I(s, \phi, U, \uparrow) \tag{28}$$

By using this spin dependence of the tunneling current in spin-polarized STM (SPSTM) experiments, magnetic information about solid surfaces can be obtained.

Spin-dependent tunneling had been observed in the 1970s using planar tunnel junctions [35–37]. To explain the experimental results, Slonczewski [38] considered a tunnel junction with two ferromagnetic electrodes where the directions of the internal magnetic fields differ by an angle θ (Fig. 10). Within a free-electron model and in the limit of a small applied bias voltage, the following expression for the conductance σ of the ferromagnet/insulator/ferromagnet tunnel junction for the case of two identical ferromagnetic electrodes was found:

$$\sigma = \sigma_{\mathrm{fbf}}(1 + P_{\mathrm{fb}}^2 \cos\theta), |P_{\mathrm{fb}}| \leqslant 1 \tag{29}$$

Here, P_{fb} denotes the effective spin polarization of the ferromagnet/barrier interface

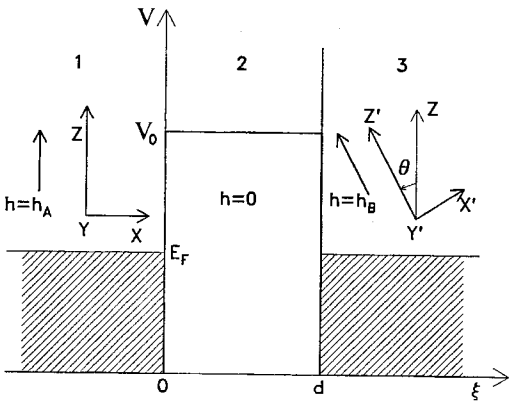


Figure 10. Schematic potential diagram for two metallic ferromagnets separated by an insulating barrier. The molecular fields h_A and h_B within the magnets form an angle θ. (Image courtesy of Slonczewski [38].)

and σ_{fbf} is a mean conductance which is proportional to $\exp(-2\kappa s)$. If the ferromagnetic electrodes are different, the conductance becomes

$$\sigma = \sigma_{fbf'}(1 + P_{fb}P_{f'b}\cos\theta) \tag{30}$$

For the two special cases of parallel and antiparallel alignment of the internal magnetic field directions, one finds

$$\begin{aligned}\sigma_{\uparrow\uparrow} &= \sigma_{fbf'}(1 + P_{fb}P_{f'b}) \\ \sigma_{\uparrow\downarrow} &= \sigma_{fbf'}(1 - P_{fb}P_{f'b})\end{aligned} \tag{31}$$

Consequently, one obtains

$$\frac{\sigma_{\uparrow\uparrow} - \sigma_{\uparrow\downarrow}}{\sigma_{\uparrow\uparrow} + \sigma_{\uparrow\downarrow}} = P_{fb}P_{f'b} = P_{fbf'} \tag{32}$$

where $P_{fbf'}$ is the effective polarization for the whole tunnel junction. An experimental determination of the quantity on the left-hand side of Eq. (32) by means of SPSTM offers a way to derive the effective polarization $P_{fbf'}$ locally with a spatial resolution comparable to that of topographic STM images and therefore ultimately on the atomic scale.

The spin dependence of the tunneling current in SPSTM experiments with two magnetic electrodes in a zero external magnetic field was demonstrated by using a ferromagnetic CrO_2 probe tip and a Cr(001) surface [39]. The topological antiferromagnetic order of the Cr(001) surface [40] with alternately magnetized terraces separated by monoatomic steps was confirmed. In addition, a local effective polarization of the CrO_2/vacuum/Cr(001) tunnel junction was derived. Later, atomic resolution in SPSTM experiments has been demonstrated on a magnetite (Fe_3O_4) (001) surface, where the two different magnetic ions Fe^{2+} and Fe^{3+} on the Fe B-sites in the Fe–O (001) planes could be distinguished by using an atomically sharp Fe probe tip prepared in situ [41].

With an additional external magnetic field applied, the magnetization of the sample (or of the tip) can be modulated periodically, for instance, from parallel to antiparallel alignment relative to the tip (or the sample) magnetization direction. Consequently, a portion of the tunneling current is predicted to oscillate at the same frequency, with an amplitude linearly proportional to the average tunneling current [42]. The advantage of this experimental procedure lies in the fact that lock-in detection techniques can be used, resulting in an improvement of the signal-to-noise ratio. In principle, the magnetic field can be modulated at a frequency ν_0 well above the cut-off frequency of the feedback loop, and the corresponding amplitude of the current oscillation at the frequency ν_0 can be recorded with a lock-in amplifier simultaneously with the constant-current topograph. The spatially resolved lock-in signal then provides a map of the effective spin polarization.

A third approach to SPSTM is to use GaAs either as samples [43, 44] or as tips. It is well known that GaAs optically pumped by circularly polarized light provides an efficient source for spin-polarized electrons. On the other hand, one can measure the circular polarization of the recombination luminescence light induced by electrons tunneling from a ferromagnetic counterelectrode.

1.6 Inelastic Tunneling Spectroscopy

Besides elastic tunneling processes, in which the energy of the tunneling electrons is conserved, inelastic tunneling can occur where the electron energy is changed due to interaction of the tunneling electrons with elementary excitations. In Fig. 11 a potential energy diagram is shown, illustrating elastic and inelastic tunneling processes. In the case of inelastic tunneling, the electron loses a quantum of energy $h\nu_0$ to some elementary excitation mode. According to the Pauli exclusion principle, the final state after the inelastic tunneling event must be initially unoccupied as depicted in Fig. 11a. Consequently, the bias dependence of the tunneling current (Fig. 11b) shows the behavior described below.

Starting from a zero applied bias voltage U, the elastic tunneling current increases linearly proportional to U

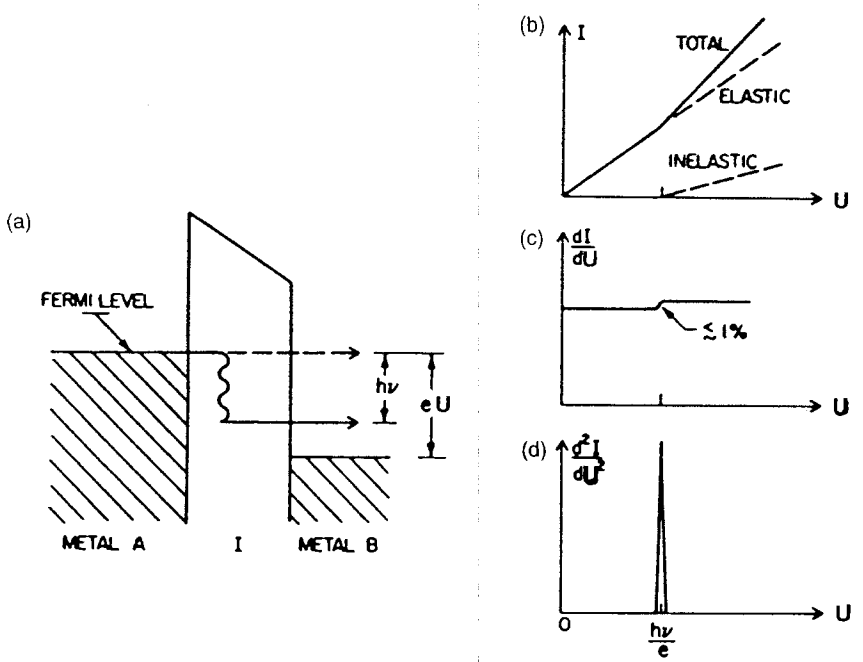


Figure 11. (a) Tunneling electrons can excite a molecular vibration of energy $h\nu$ only if $eU > h\nu$. For smaller voltages there are no vacant final states for the electrons to tunnel into. (b) Thus the inelastic current has a threshold at $U = h\nu/e$. (c) The increase in conductance at this threshold is typically below 1%. (d) A standard tunneling spectrum, d^2I/dU^2 versus U, accentuates this small increase; the step in dI/dU becomes a peak in d^2I/dU^2. (Image courtesy of Hansma [45].)

(Eq. (4)). As long as the applied bias voltage is sufficiently small ($U < h\nu_0/e$, where ν_0 is the lowest energy excitation mode), inelastic tunneling processes cannot occur due to the Pauli exclusion principle. At the threshold bias $U_0 = h\nu_0/e$, the inelastic channel opens up, and the number of electrons which can use the inelastic channel will increase linearly with U (Fig. 11b). Therefore, the total current, including both elastic and inelastic contributions, has a kink at $U_0 = h\nu_0/e$. In the conductance (dI/dU) versus voltage curve, the kink becomes a step at U_0. Since the fraction of electrons which tunnel inelastically is tiny (typically 0.1–1%), the conductance increase at U_0 due to the onset of the inelastic tunnel channel is too small to be conveniently observed. Therefore, the second derivative (d^2I/dU^2) is usually measured, which exhibits a peak at U_0. In general, there are many modes which can be excited in the tunneling process. Each excitation mode ν_i contributes a peak in the second derivative $d^2I/dU^2(U)$ at the corresponding bias voltage $U_i = h\nu_i/e$ so that $d^2I/dU^2(U)$ represents the spectrum of possible excitations. Inelastic electron tunneling can therefore be regarded as a special kind of electron energy loss spectroscopy.

To be able to detect the small changes in tunneling conductance $\Delta G/G$ as a result of the opening of additional inelastic tunneling channels, the relative stability of the tunneling current has to be better than 1%. In addition, low temperatures are required to keep thermal linewidth broadening in the spectra, which is of the order of $k_B T$, small compared with the energy $h\nu$ of the excitation modes, ν being typically a few millielectronvolts.

1.6.1 Phonon Spectroscopy

Electron tunneling between the probe tip and the sample can create phonons at the interface between the conductor and the tunneling barrier. The emission of phonons is believed to take place within a few atomic layers of the interface. Low-temperature STM experiments with a tungsten probe tip and a graphite sample indeed revealed a spectrum of peaks in $d^2I/dU^2(U)$ characteristics where the positions of the peaks corresponded closely to the energies of the phonons of the graphite sample and the tungsten tip [46]. The measured increase in conductance at the phonon energies was of the order of 5%. By analogy with elastic scanning tunneling spectroscopy (see Sec. 1.4.1 of this Chapter), spectroscopic imaging can be performed by recording d^2I/dU^2 at a particular phonon energy while scanning the tip over the sample surface. This method allows one to map spatial variations of the phonon spectra, caused by spatial variations in the coupling between the tunneling electrons and the phonons, on the atomic scale.

1.6.2 Molecular Vibrational Spectroscopy

Inelastic tunneling spectroscopy can also yield information about vibrational modes of molecules adsorbed on a surface. By using low-temperature STM, a vibrational spectrum of an individual adsorbed molecule can be obtained by positioning the probe tip over the preselected adsorbate. It is even possible to form a map showing the sites within a molecule where particular resonances occur.

For sorbic acid adsorbed on graphite, a spectrum of strong peaks was observed in the first derivative dI/dU instead of the expected second derivative d^2I/dU^2 [47]. The energies of the peaks corresponded approximately to the vibrational modes of the molecule. The measured increase in conductivity at the molecular resonances was as much as a factor of 10, which is at least two orders of magnitude larger than expected.

Future inelastic tunneling experiments have to focus on the assignment of characteristic features in the tunneling spectra to particular molecular functional groups. This will probably help to identify chemical species by STM, a problem which is not solvable by elastic tunneling spectroscopy.

Acknowledgments

The author would like to acknowledge all colleagues who provided illustrations for this Chapter.

1.7 References

[1] G. Binnig, H. Rohrer, Ch. Gerber, E. Weibel, *Phys. Rev. Lett.* **1982,** *49*, 57.
[2] G. Binnig and H. Rohrer, *Helv. Phys. Acta* **1982,** *55*, 726.
[3] G. Binnig and H. Rohrer, *Rev. Mod. Phys.* **1987,** *59*, 615.
[4] R. Wiesendanger, *Scanning Probe Microscopy and Spectroscopy: Methods and Applications*, Cambridge University Press, Cambridge **1994**.
[5] J. Bardeen, *Phys. Rev. Lett.* **1961,** *6*, 57.
[6] J. Tersoff, D. R. Hamann, *Phys. Rev. Lett.* **1983,** *50*, 1988.
[7] J. Tersoff, D. R. Hamann, *Phys. Rev. B* **1985,** *31*, 805.
[8] C. J. Chen, *J. Vac. Sci. Technol.* **1988,** *A6*, 319.
[9] S. Ohnishi, M. Tsukada, *J. Vac. Sci. Technol.* **1990,** *A8*, 174.
[10] C. J. Chen, *Phys. Rev. B* **1990,** *42*, 8841.
[11] C. J. Chen, *J. Vac. Sci. Technol.* **1991,** *A9*, 44.
[12] N. D. Lang, *Phys. Rev. Lett.* **1985,** *55*, 230.
[13] N. D. Lang, *Phys. Rev. Lett.* **1986,** *56*, 1164.
[14] N. D. Lang, *IBM J. Res. Develop.* **1986,** *30*, 374.
[15] Y. Kuk, P. J. Silverman, H. Q. Nguyen, *J. Vac. Sci. Technol.* **1988,** *A6*, 524.
[16] J. Wintterlin, J. Wiechers, H. Brune, T. Gritsch, H. Höfer, R. J. Behm, *Phys. Rev. Lett.* **1989,** *62*, 59.
[17] M. Bode, R. Pascal, R. Wiesendanger, *Z. Phys. B*, in press.
[18] V. M. Hallmark, S. Chiang, J. F. Rabolt, J. D. Swalen, R. J. Wilson, *Phys. Rev. Lett.* **1987,** *59*, 2879.
[19] A. Baratoff, *Physica B* **1984,** *127*, 143.
[20] C. J. Chen, *Phys. Rev. Lett.* **1990,** *65*, 448.
[21] G. Binnig, N. Garcia, H. Rohrer, J. M. Soler, F. Flores, *Phys. Rev. B* **1984,** *30*, 4816.
[22] J. H. Coombs, M. E. Welland, J. B. Pethica, *Surf. Sci.* **1988,** *198*, L353.
[23] G. Binnig, H. Rohrer, *Surf. Sci.* **1983,** *126*, 236.
[24] R. J. Hamers, *Annu. Rev. Phys. Chem.* **1989,** *40*, 531.
[25] R. M. Tromp, *J. Phys. C: Condens. Matter* **1989,** *1*, 10211.
[26] R. M. Feenstra, *NATO ASI Ser. E: Appl. Sci.* **1990,** *184*, 211.
[27] T. Klitsner, R. S. Becker, J. S. Vickers, *Phys. Rev. B* **1990,** *41*, 3837.
[28] G. Binnig, K. H. Frank, H. Fuchs, N. Garcia, B. Reihl, H. Rohrer, F. Salvan, A. R. Williams, *Phys. Rev. Lett.* **1985,** *55*, 991.
[29] R. S. Becker, J. A. Golovchenko, D. R. Hamann, B. S. Swartzentruber, *Phys. Rev. Lett.* **1985,** *55*, 2032.
[30] R. M. Feenstra, W. A. Thompson, A. P. Fein, *Phys. Rev. Lett.* **1986,** *56*, 608.
[31] J. S. Stroscio, R. M. Feenstra, A. P. Fein, *Phys. Rev. Lett.* **1986,** *57*, 2579.
[32] R. M. Feenstra, J. A. Stroscio, *Phys. Scripta* **1987,** *T19*, 55.
[33] R. J. Hamers, R. M. Tromp, J. E. Demuth, *Phys. Rev. Lett.* **1986,** *56*, 1972.
[34] P. Avouris, I.-W. Lyo in *Chemistry and Physics of Solid Surfaces VIII. Springer Series in Surface Sciences*, Vol. 22 (Eds.: R. Vanselow, R. Howe), Springer, Berlin, **1990,** p. 371.
[35] P. M. Tedrow, R. Meservey, *Phys. Rev. Lett.* **1971,** *26*, 192.
[36] M. Julliere, *Phys. Lett. A* **1975,** *54*, 225.
[37] S. Maekawa, U. Gäfvert, *IEEE Trans. Magn.* **1982,** *18*, 707.
[38] J. C. Slonczewski, *Phys. Rev. B* **1989,** *39*, 6995.
[39] R. Wiesendanger, H.-J. Güntherodt, G. Güntherodt, R. J. Gambino, R. Ruf, *Phys. Rev. Lett.* **1990,** *65*, 247.

[40] S. Blügel, D. Pescia, P. H. Dederichs, *Phys. Rev. B* **1989,** *39*, 1392.
[41] R. Wiesendanger, I. V. Shvets, D. Bürgler, G. Tarrach, H.-J. Güntherodt, J. M. D. Coey, *Europhys. Lett.* **1992,** *19*, 141.
[42] M. Johnson, J. Clark, *J. Appl. Phys.* **1990,** *67*, 6141.
[43] S. F. Alvarado, P. Renaud, *Phys. Rev. Lett.* **1992,** *68*, 1387.
[44] K. Sueoka, K. Mukasa, K. Hayakawa, *Jpn. J. Appl. Phys.* **1993,** *32*, 2989.
[45] P. K. Hansma, *Phys. Rep.* **1977,** *30*, 145.
[46] D. P. E. Smith, G. Binnig, C. F. Quate, *Appl. Phys. Lett.* **1986,** *49*, 1641.
[47] D. P. E. Smith, M. D. Kirk, C. F. Quate, *J. Chem. Phys.* **1987,** *86*, 6034.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

2 Scanning Force Microscopy

2.1 Introduction

The invention of scanning tunneling microscopy (STM) in 1982 [1] triggered the development of several further techniques which use scanned point probes to sense local properties of surfaces [2]. Among these techniques, scanning force microscopy (SFM), which was originally denoted atomic force microscopy (AFM) [3], has developed into the most widespread and commercially successful tool, and is used not only in physical, chemical, biological, and medical research laboratories, but also by many companies for tasks such as product development and routine quality control.

The history of SFM started in the fall of 1985, when Gerd Binnig and Christopher Gerber, on leave from IBM's research laboratory in Zurich, Switzerland, built the prototype of a force microscope together with Calvin Quate at his group at Stanford University, California, U.S.A. [3]. The success of SFM in the following years was due to the high resolution and the versatility of this new tool, which can map not only the pure topography of sample surfaces from the micrometer to the atomic scale, but also (often additionally to the simultaneously recorded topography) many other physical properties which are related to forces. Moreover, due to its universality, SFM can be applied to a large variety of samples. It requires no special sample preparation and can be adapted to many different environments such as air, vacuum, and liquids. These issues are exemplified in detail below.

The principle of SFM is rather simple, and is analogous to that of a record player. A force microscope (see Fig. 1) detects forces acting between a sample surface and a sharp tip which is mounted on a soft leaf spring, the so-called cantilever. A feedback system which controls the vertical z-position of the tip on the sample surface keeps the deflection of the cantilever (and thus the force between the tip and sample) constant. Moving the tip relative to the sample in the (x, y) plane of the surface by means of piezoelectric drives, the actual z-position of the tip can be recorded as a function of the lateral (x, y) position. The obtained three-dimensional data represent a map of equal forces. The data can be analyzed and visualized through computer processing.

With the concept described above, it is possible to detect any kind of force as long as the tip is sensitive enough to the interaction, that is, the interaction causes a measurable deflection of the cantilever on which the tip is mounted. Therefore, not only interatomic forces

but also long-range forces such as magnetic or electrostatic forces can be detected and mapped. In order to manifest the general concept, the original name 'atomic force microscopy' [3] was replaced by the term 'scanning force microscopy', which is used in this chapter. In a typical force microscope, cantilever deflections in the range from 0.1 Å to a few micrometers are measured [4–6], corresponding to forces from 10^{-13} to 10^{-5} N. For comparison, typical forces in conventional profilometers are 10^{-4} N or greater [7].

Force microscopes reach an impressive magnification. Images of atomic lattices have been demonstrated with a lateral resolution well below 1 Å. The maximum scan range is usually limited by the choice of the piezoelectric scanner. Most manufacturers supply scanners up to at least the 100×100 μm range. Even if large areas are imaged, the resolution in the z-direction is still sufficient to image monoatomic steps (cf. Fig. 4b).

There are only few limitations in the type of sample which can be investigated by SFM. A sample is suitable to be studied by SFM as long as it is solid and clean. For example, photographic material, which is an insulator and highly sensitive to light or electron beams, can be easily studied without any surface damage [8]. In a liquid environment, even soft samples such as the biological membrane of a virus or a cell have been successfully imaged [9, 10]. However, if powder particles [11], grains from colloidal solution [12], single molecules [13], or similar samples are to be investigated, they should adhere to a substrate by forces stronger than the lateral forces of the scanning tip. This is often realized by physisorption [14] or chemisorption [15].

The possible *sample size* is, in principle, *unlimited*, since stand-alone microscopes can be put on the sample itself [16]. Special large-sample microscopes are sold for routine quality control of whole wafers in the electronic industry at different steps of fabrication. After the spot check, the wafer can be reintroduced into the production process even if sensitive devices are being produced, since the sample is not affected by the measurement.

Probably the most important limiting factor for the successful operation of SFM on a hard solid sample is the cleanliness of the surface which is to be imaged. Images on the micrometer scale can be obtained on nearly every sample unless they show extraordinary adhesion due to, for example, surfactants or thick adsorbent films. For high-resolution SFM micrographs, however, the absence of adsorbents which disturb the correct profiling of the surface is essential. Crystals which are to be imaged on the atomic scale should preferably be cleaved before the measurements.

Samples can be probed without any special treatment and in many different environmental conditions, such as in ambient air, a large variety of liquids [17, 18], gaseous atmospheres such as dry nitrogen or argon [19–21], and ultrahigh vacuum (UHV) [4, 22–24]. Measurements can be performed at different temperatures starting from temperatures close to absolute zero [25] up to temperatures well above room temperature. Additionally, SFM can easily be combined in situ with optical microscopy.

Section 2.2 of this Chapter discusses experimental aspects such as the properties of force sensors, different experimental setups, and possible imaging modes. Some

theoretical aspects are considered in Sec. 2.3 of this Chapter. Further general information about SFM can be found in reviews [26] and textbooks [27].

2.2 Experimental Aspects

Figure 1 shows a typical set-up of a force microscope. Essential components of every force microscope are (1) a local force probe (a sharp tip) which is mounted on a soft spring (the cantilever), (2) a sensor for the accurate measurement of the cantilever deflection, (3) a device for the relative motion of the tip and sample (usually a piezoelectric scanner), (4) a feedback system to control the cantilever deflection, and (5) computer-based electronics for the visualization and analysis of the measured data. Components (1)–(3) are assembled in a rigid mechanical set-up. Components (3)–(5) are identical in principle to the corresponding devices used in STM, and will not be described here. General design criteria for scanning probe microscopes are described in detail in the literature [28].

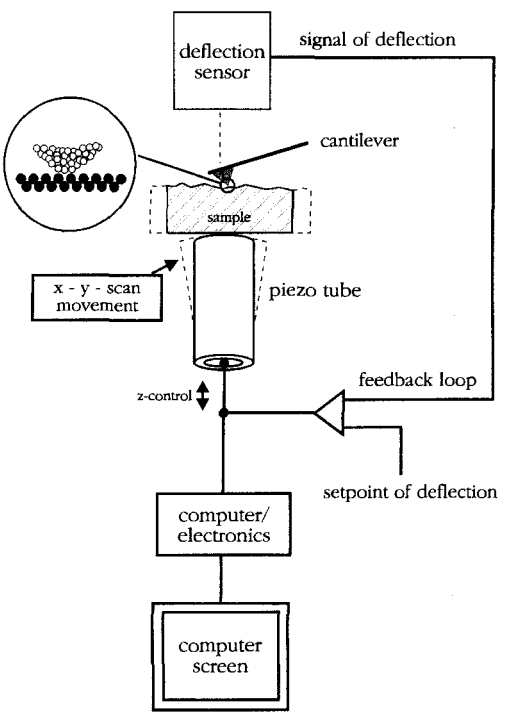


Figure 1. Schematic of the set-up of a typical force microscope operating in the constant force imaging mode.

However, since the principle of SFM is based on the measurement of the forces acting between a sharp tip and the sample, the preparation of suitable tips and cantilevers is of pre-eminent importance, and will be discussed in Sec. 2.2.1 of this Chapter. Even small deflections of the cantilever have to be detected (see Sec. 2.2.2). Finally, force microscopes can be operated in different modes, which are analyzed in Secs. 2.2.3–2.2.6 of this Chapter.

2.2.1 The Force Sensor

In force microscopy, both the spring and the force probe (the sharp tip), which is mounted on the spring, have to satisfy certain requirements. There are basically two requirements for the spring:

(1) The transmission of an external vibration to the system is described by $a_{\text{trans}} = a_0(f_0/f_{\text{res}})^2$, where f_0 is a frequency of the vibration amplitude a_0 and f_{res} is the resonance frequency of the spring. It is usual to choose cantilevers with a high value of f_{res} to omit problems due to acoustic or mechanical vibrations of low frequency. This means, since f_{res} is

proportional to $\sqrt{c/m}$ (c = spring constant, m = mass of the spring), that the ratio c/m has to be large.

(2) The spring constant c should be small to keep the forces acting on the sample surface as small as possible. Ideally, it should be well below interatomic spring constants, which are of the order of $10\,\mathrm{N\,m^{-1}}$.

Conditions (1) and (2) can only be fulfilled simultaneously if the mass m of the spring is small. Today, springs of a small mass are produced by standard microfabrication techniques from silicon or silicon nitride in the form of rectangular or 'V'-shaped cantilevers (Fig. 2a) [29]. The typical dimensions of such cantilevers are as follows: lengths of some hundreds of micrometers, widths of some tens of micrometers, and thicknesses of 0.3–5 μm. Since the dimensions are within a quite narrow range, the spring constants can be calculated accurately. Spring constants and resonance frequencies of cantilevers used in contact SFM measurements are about $0.01–1\,\mathrm{N\,m^{-1}}$ and 5–100 kHz, respectively.

A tip which acts as a local probe is mounted at the end of the cantilever (Fig. 2b,c). The first tips were simply small pieces of diamond glued to the end of cantilevers which were cut from metal foil [3]. Later, microfabrication techniques were used to produce the cantilevers, to which the diamonds were glued [29]. Sometimes, thin metal wires were bent and etched at their ends. This kind of tip was frequently used in magnetic force microscopy. During recent years, however, new techniques have been developed to produce microfabricated cantilevers with integrated tips of high quality [30].

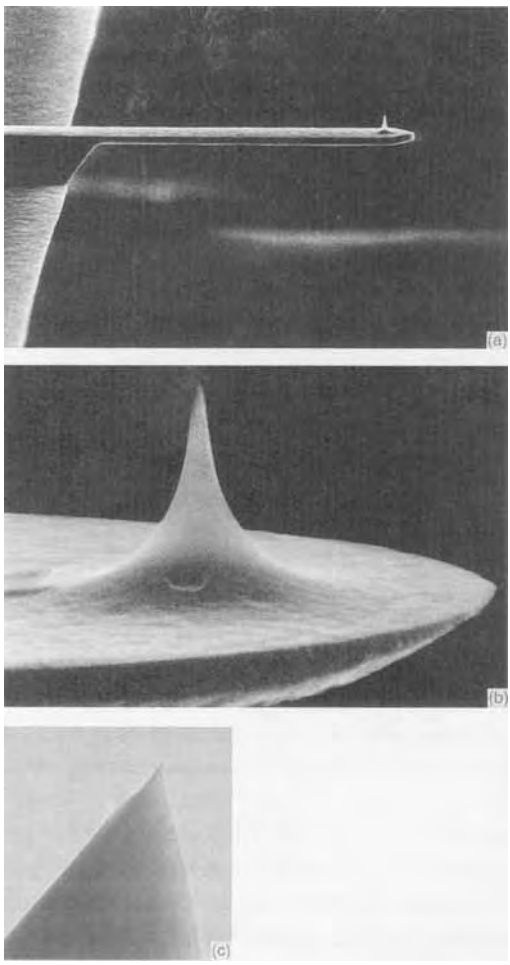


Figure 2. (a) Scanning electron micrograph of a rectangular silicon cantilever of 105 μm length and 14.5 μm width. (b) Scanning electron micrograph of the tip at the end of the cantilever shown in part (a). The tip height is 6.5 μm. (c) Transmission electron micrograph of the tip end of a silicon cantilever. The image size is 2 × 2 μm, and the tip radius is estimated to be below 10 nm.

The shape of an SFM is often approximated by a cone which has a small half-sphere at its end. The tip should be as sharp as possible to measure very local properties, that is, it should possess a small opening angle for the cone and a

small radius for the sphere. On the other hand, tips with a small opening angle are often unstable and tend to flex or break. Silicon nitride tips are more robust than silicon tips due to the greater hardness of silicon nitride. However, silicon nitride tips are currently not available with such small opening angles and tip radii as silicon tips. Therefore, it is important to choose a suitable tip for a specific application [31].

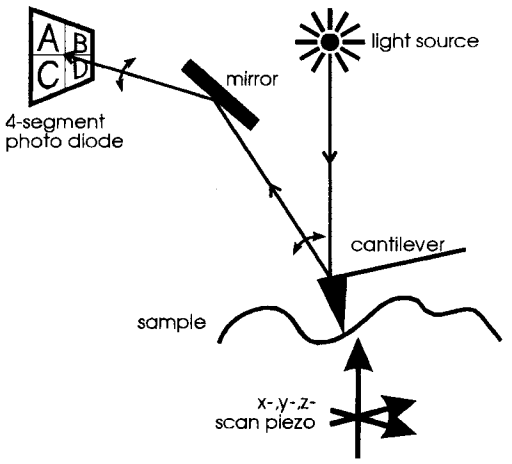


Figure 3. Schematic drawing of the beam deflection technique for the sensing of cantilever deflections in force microscopy. The light from a light source is reflected at the back face of the cantilever and focused onto a four-segment photodiode. By analyzing the photocurrents induced in the different segments A–D, the deflection as well as the torsion of the cantilever can be detected simultaneously.

2.2.2 Deflection Sensors

Several techniques have been developed to detect the deflection of the cantilever, which has to be measured with sub-angstrom resolution. The most frequently used techniques are reviewed in this section.

The beam deflection technique was developed independently by Meyer and Amer [4] and Alexander et al. [5] in 1988, and is currently the most widespread technique used to measure cantilever deflections in force microscopy. The idea behind this technique is presented in Fig. 3. A light beam from a laser diode or a high-power light-emitting diode is reflected from the rear side of the cantilever and focused onto a four-segment photodiode. If I_A, I_B, I_C, and I_D are the currents which are induced by the light in segments A–D of the photodiode, then the current $(I_A + I_B) - (I_C + I_D)$ represents a measure of the deflection of the cantilever. Additionally, the torsion of the cantilever can be measured through the analysis of the $(I_A + I_C) - (I_B + I_D)$ current, which is proportional to the lateral force acting on the tip (cf. Sec. 2.2.3.4 of this Chapter) [32]. This technique is used in most commercially available force microscopes, mainly due to a high z-resolution (typically of 0.1 Å) and the easy to measure torsion of the cantilever.

Interferometric detection schemes were introduced even before the beam deflection technique was presented [33, 34]. The idea underlying this detection scheme is explained below. A laser beam, reflected from the rear of the cantilever, interferes with a reference beam, the changes in intensity of the signal are proportional to the deflection of the cantilever. Today, most of the interferometers used to sense cantilever deflection in SFM are based on glass fiber technology [6]. This allows the positioning of the light source and the photodetector at a distance from the force microscope, enabling a compact microscope design [35]. Furthermore, miniaturized devices with dimensions of

only a few millimeters are under development. The resolution which can be obtained with this technique is similar to the beam deflection technique [36].

An alternative optical method is laser diode interferometry [37]. The light from a laser diode is reflected from the rear of the cantilever back into the laser cavity. Laser diodes are extremely sensitive to this type of feedback, which causes measurable changes in the laser intensity. In spite of the simplicity of the set-up, the analysis of the signal is rather complex [38].

A comparatively new approach is the use of piezoresistive cantilevers, which serve as deflection sensors [39]. A piezoresistive film at the rear of the cantilever changes its resistance if the cantilever is bent. The change of resistance is proportional to the deflection, and is detected by a Wheatstone bridge. The main advantage of this technique is that the size of the force microscope can be reduced since an optical part is not necessary. However, the big disadvantage of this technique is the still insufficient sensitivity when combined with soft cantilevers for contact measurements. Therefore, such microscopes are mostly operated in the non-contact dynamic mode (see Sec. 2.2.3.3 of this Chapter) [24, 40].

Historically, the first techniques were the electron tunneling technique and the capacitance technique. The electron tunneling technique was applied in the first force microscope of Binnig, Gerber, and Quate [3] and in several other early force microscopes [41]. In these the deflection of the cantilever was sensed by a tunneling junction between the back of the cantilever and an additional tip. It showed an excellent z-resolution due to the exponential dependence of the tunneling current on the separation of the electrodes (see Sec. 1 of this Chapter). Many experimental difficulties such as the instability of the tunneling junction and a sensitivity to the surface roughness of the rear of the cantilever led to the abandonment of this technique.

Force microscopes equipped with capacitance sensors do not yet show a performance comparable to other methods [23, 42]. In these microscopes, the change in the capacitance between the rear side of the cantilever and an external electrode is measured. In the future, microfabricated capacitances which are integrated into the cantilever design might significantly improve the reliability and sensitivity of this method.

2.2.3 Imaging Modes

Different imaging scenarios are realized depending on which physical parameter is taken to the feedback system. The speed of the feedback system and the number of parameters which are monitored and recorded determine the imaging operation mode. A survey of the most common modes of SFM operation is given below.

2.2.3.1 Constant Force Mode

This is the standard method of SFM operation. The cantilever deflection in the z-direction (and hence the force acting between the tip and the sample) is kept constant using a feedback loop (see Fig. 1). Thus, if the sample is scanned in the (x, y) plane, the output of the feedback gives a map of equal force which usually corresponds to the topography of the sample. Some SFM micrographs acquired in this mode are presented in Fig. 4.

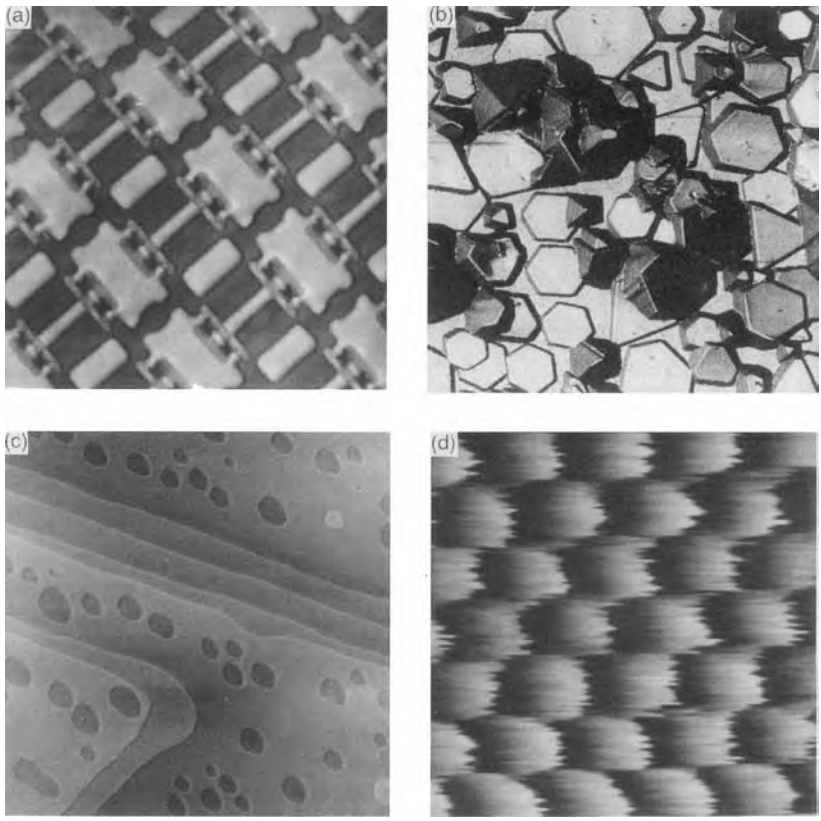


Figure 4. Examples of SFM micrographs acquired in the constant force mode, demonstrating possible applications of SFM in science and technology. (a) A $70 \times 70\,\mu m$ area of an integrated circuit. Individual components are visible. Light bright regions represent high surface areas and dark regions low surface areas. (b) Tabular silver bromide microcrystals (so-called 'T-grains') as used in modern photographic material, deposited on a glass substrate. The scanned area was $30 \times 30\,\mu m$; individual microcrystals are 100–400 nm high. The image is displayed as a stereoscopic projection (bird's-eye view). Despite the high sensitivity of the material to visible light or electronic beams, the microcrystals can be imaged without surface damage [8]. (c) A $5 \times 5\,\mu m$ area of a (010) cleavage face of triglycine sulfate. Light regions represent high surface areas and dark regions low surface areas. The step height between individual terraces is 6 Å, which corresponds to half of the unit cell of the material. (d) High-resolution SFM micrograph of a mica surface, demonstrating molecular resolution. The regular protrusions have the periodicity of the lattice (5.2 Å distance).

2.2.3.2 Variable Deflection Mode

If the sample is scanned with respect to the cantilever without any feedback, an image of the surface is obtained by the direct recording of the output of the deflection sensor. Higher scan rates (to nearly video frequencies) can be achieved in this mode. This is an advantage when imaging on the atomic scale. Large scans, however, might lead to a deformation or even a destruction of the tip and/or sample due to changing force between the tip and the sample (see Sec. 2.2.6 of this Chapter).

2.2.3.3 Noncontact Dynamic Modes

The resonance frequency f_{res} of the cantilever is proportional to $\sqrt{c - \partial F/\partial z}$, where c is the spring constant of the cantilever and $\partial F/\partial z$ is the gradient of the tip–sample interaction force normal to the sample surface. A change in the force gradient results in a change of the resonance frequency which is measured [33, 34]. An example of an experimental set-up is sketched in Fig. 5. The cantilever is vibrated close to its resonance frequency by a piezo element. Then, the signal obtained by the deflection sensor is analyzed by the lock-in technique. A subsequent feedback circuit regulates on a constant phase shift between the signal of the deflection sensor and the original driving signal or on a constant amplitude of the modulated deflection sensor signal (cf. Fig. 5 in Sec. 3 of this Chapter). Both methods keep the resonance frequency f_{res} constant.

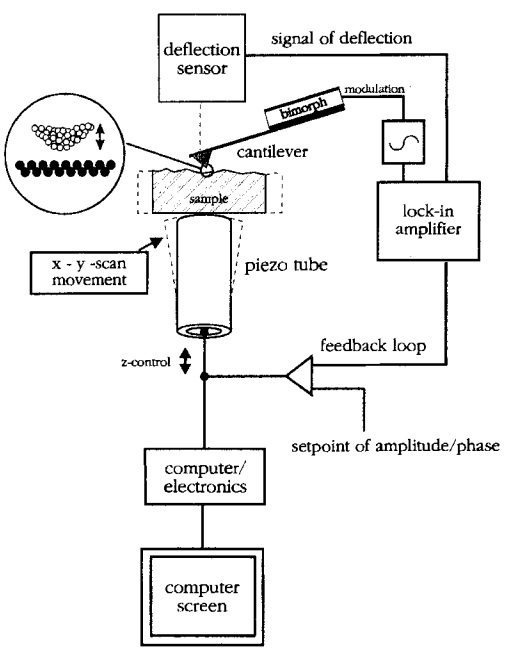


Figure 5. Schematic of a typical set-up of a force microscope operated in the noncontact dynamic imaging mode.

The noncontact dynamic mode shows important differences in comparison to the imaging modes described above. First, since the resonance frequency f_{res} and not the deflection of the cantilever is kept constant, lines of constant force gradient (and not of constant force) are recorded. Second, the tip does not touch the sample during the measurement. Therefore, surface deformations and lateral forces are minimized [10, 43]. Finally, since a surface can be traced not only for a distance of a few nanometers, but also for tens of nanometers or even more, long-range forces such as electrostatic [43, 44] or magnetic (see Sec. 3.2 of this Chapter) forces can be imaged separated from the surface topography.

In order to obtain a maximum signal-to-noise ratio, the cantilever should possess a high resonance frequency f_{res} and a low force constant c [33]. However, the thermally induced motion of the cantilever, which is proportional to $c^{-1/2}$, reduces the minimal detectable value of a force gradient [33]. Furthermore, the lateral resolution increases with decreasing distance between the tip and the sample. In order to obtain the highest resolution, the tip must profile the sample surface as close as possible. If the spring constant is smaller than the local force gradient normal to the sample surface, the cantilever snaps to the surface (see Sec. 2.2.4 of this Chapter). This happens with soft springs which are suitable for contact force measurement (force constants between 0.01 and $1\,N\,m^{-1}$). Therefore, hard cantilevers with spring constants between $5\,N\,m^{-1}$

and more than $100\,N\,m^{-1}$ (and corresponding resonance frequencies of some 100 kHz) are usually chosen to scan close to the surface. Then, oscillation amplitudes have to be restricted to a few angstroms.

If the damping of the cantilever oscillation is very low (i.e., the quality Q of the cantilever is very high), it is difficult to regulate on a constant amplitude or constant phase shift. Therefore, Albrecht et al. [45] suggested a method for the direct determination of the cantilever resonance frequency which is especially useful for vacuum measurements. A detailed theoretical description of noncontact force microscopy is given by Hartmann [46].

2.2.3.4 Imaging Friction, Elasticity, and Viscosity

In the so-called lateral force microscopy (LFM), which is often also termed friction force microscopy (FFM), the torsion of the cantilever is measured additionally to the simultaneously recorded topography [47]. Experimentally, this is usually realized by using the beam deflection technique described in Sec. 2.2.2 of this Chapter. A map of the lateral forces acting on the tip is generated. In the absence of topographical effects (topographical steps or slopes cause a torque of the tip [47, 48]), the torsion is proportional to the frictional force between the tip and the sample. Variations of the frictional force on the atomic level have been observed [49]. Figure 6a shows the topography and Fig. 6b the simultaneously recorded friction force map of a thin film of C_{60} molecules on a GeS substrate. The C_{60} islands (dark areas in Fig. 6b) exhibit a lower friction than the GeS substrate.

The viscoelastic properties of samples can be investigated by the modulation of the vertical sample position [50]. The modulation of the sample position (or, alternatively, the position of the cantilever support) leads to a modulation of the force acting between the tip and the sample. Using the lock-in technique, the in-phase

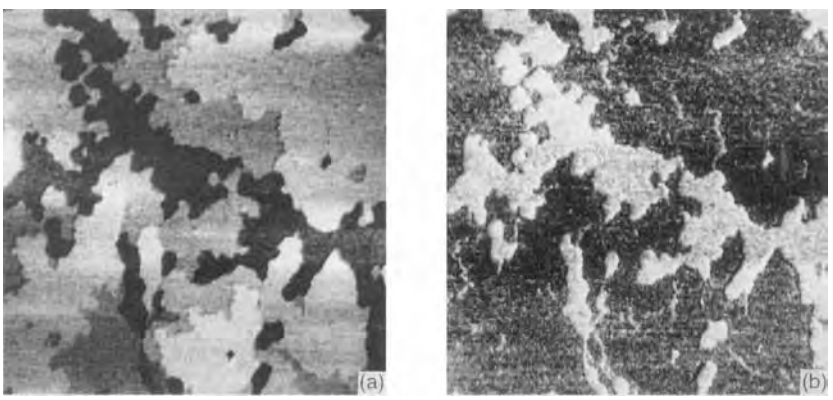


Figure 6. (a) Topography and (b) simultaneously recorded friction force map of the thin film of C_{60} molecules epitaxially grown on a GeS substrate. The scanned area is $2 \times 2\,\mu m$. The C_{60} islands are one or two monolayers high; each monolayer corresponds to about 1 nm in height. The C_{60} islands (dark regions in part (b)) exhibit a lower friction than the GeS substrate (light regions).

amplitude and the phase shift of the response of the sample can be recorded, which correspond to the elasticity and the viscosity of the surface.

2.2.3.5 Other Imaging Modes

Several further imaging modes or combinations of imaging modes are possible such as the so-called 'lift mode' (see Sec. 3.1 of this Chapter), which combines topography measurement and measurement of a long-range force. Another example is dynamic high-amplitude measurement ('tapping mode'), where the cantilever is vibrated as described for the noncontact modes, but purposely touches the surface at each cycle [51].

2.2.4 Force–Distance Curves

Since the force microscope probes the force between a tip and a sample, SFM can be used to study the tip–sample interaction as a function of their separation. The force–distance (fd) curve provides useful information about both long- and short-range forces as well as surface hardness, etc.

Figure 7 shows a typical fd curve. The horizontal axis is calibrated so that the sample position is zero if the nondeflected cantilever touches the surface. The left vertical axis shows the deflection of the cantilever. Negative values indicate a bending of the cantilever toward the sample. The loading force F (right vertical axis) can be calculated from Hooke's law $F = -cz$, where c is the spring constant ($0.1\ \mathrm{N\,m^{-1}}$ for the cantilever used in Fig. 7) and z is the deflection of the cantilever.

At position 1, the tip is distant from the surface, and no interaction occurs (cantilever not deflected). When the cantilever approaches the surface, a slight deflection of the cantilever is observed starting at a distance of about 75 nm from the surface due to the long-range van der Waals forces. At position 2, close to the sample surface, the force gradient of the attractive interaction force is larger than the spring constant of the cantilever. Thus, the attractive forces acting on the tip cannot

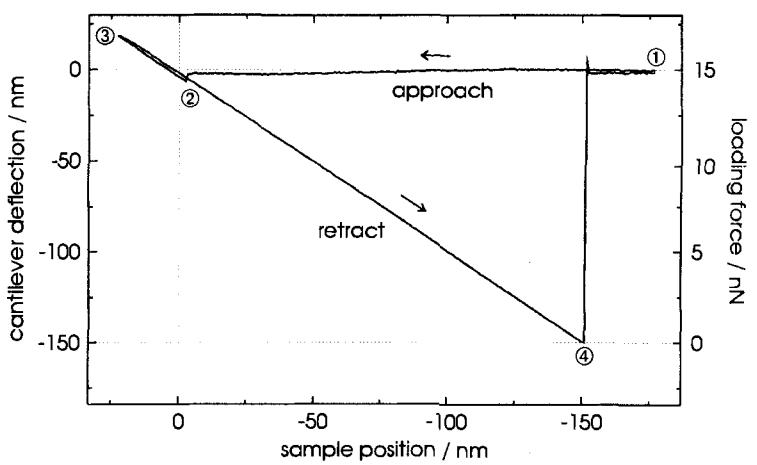


Figure 7. Cantilever deflection and loading force of the tip on the sample as a function of the tip–sample separation. At position 2, the free cantilever jumps to the sample surface and sticks there until position 4 is reached.

be compensated for by the spring force of the cantilever, and the cantilever snaps to the surface (cantilever bends toward the surface by about 6 nm in the present example). The sample is retracted starting from position 3. The tip sticks to the surface until position 4 is reached due to attractive forces such as adhesion and capillary forces. At this position, just before the tip loses the contact with the sample surface, the cantilever is strongly bent toward the sample by 150 nm (see Fig. 7), corresponding to a repulsive spring force of 15 nN. If the cantilever is bent less toward the sample, the outermost tip atom is still in a repulsive force regime, causing probable local surface deformation or even surface damage, in spite of a net attractive force on the tip [52, 53]. Only at position 4 are the attractive surface forces balanced by the spring force of the cantilever, and the repulsive force of the sample surface vanishes. Therefore, this point, which represents the force experienced by the outermost atom of the tip, is set to zero force during the standard SFM force calibration.

In summary, fd curves are essential in order to calibrate the loading force of the tip on the sample. If fd curves are performed at many different surface spots, a map of local surface adhesion can be generated [54]. The nature of long-range forces can be investigated [55]. If the end of the tip is sensitized by specific molecules, the binding strength between molecules can be measured [56]. Spatially resolved measurement of adhesion with such sensitized tips leads to functional group imaging ('chemical force microscopy', see Frisbie et al. [57]). Principally, fd curves can also be used to determine surface energies and therefore to obtain information about the local chemical composition. For this purpose, tips with well-defined shapes and advanced cantilever designs (e.g., double leaf springs [58]) are necessary. However, such tips and cantilevers are not yet commercially available.

2.2.5 Tip Artefacts

The information which can be obtained by SFM strongly depends on the kind of tip. In order to obtain a high lateral resolution, the interaction between the tip and the sample has to be limited to a very small surface region. Since real tips are not ideal point probes, tip effects have to be taken into account when interpreting the acquired data. These tip effects result from the fact that different sites of the probing tip interact with the sample during the scan, leading to a convolution of sample features with the tip shape.

An extreme example is presented in Fig. 8. It shows a latex ball with a diameter of 1 μm as used for calibration purpose in electron microscopy, scanned with a pyramidal tip. Obviously, the latex ball images the pyramidal tip more than the tip images the latex ball due to the large curvature of the ball. The original spherical shape of the ball is completely hidden.

Three types of effects mainly limit the information from SFM measurements on a scale larger than the atomic scale: (1) the surface roughness might not be reflected correctly due to the finite size of the tip end; (2) the nonzero opening angle of the tip can cause artefacts at high surface steps; and (3) double tip effects might

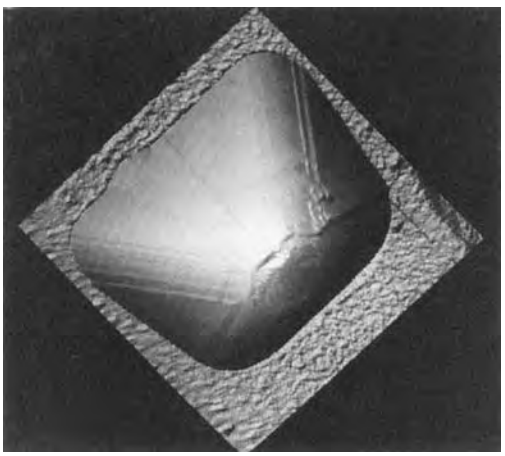


Figure 8. Spherical latex ball of 1 μm diameter on a silicon substrate, mapped with a pyramidal tip. The scan area was 2 × 2 μm. The obtained image represents a convolution of the pyramidal tip shape with the spherical ball shape.

occur on rough samples. These issues are discussed in detail by Schwarz et al. [31].

2.2.6 Scanning Force Microscopy as a Tool for Nanomodifications

If the force between tip and sample exceeds a certain limit, the sample and/or the tip is modified. The often happens accidentally when loading forces that are too large are used or on very soft materials. However, controlled increasing of the tip–sample interaction offers the possibility of performing hardness and wear tests as well as to machine surfaces or to create structures on the nanometer scale [59, 60]. The sample surface can be imaged before and after the modification with the same instrument. Additionally, suitable samples can also be structured by applying voltages and by combination of SFM with etching techniques [60, 61].

2.3 Theoretical Aspects

2.3.1 Forces in Force Microscopy

In force microscopy, a knowledge of the interactions between the tip and the sample is necessary to measure and interpret the data correctly. On the scale of atoms and molecules, the electromagnetic interaction dominates over other types of interactions. The strong and the weak interaction as well as gravitation are many magnitudes smaller. However, the electromagnetic interaction gives rise to a rich variety of different forces which not only complicate the SFM image interpretation but also have the potential of measuring many different physical properties. The potential of the force microscope lies far beyond that of a simple topography profiler. A summary of the forces relevant to force microscopy is given in Fig. 9; additional information in intermolecular and surface forces is given by Israelachvili [62].

2.3.1.1 Pauli Repulsion and Ionic Repulsion

The most important forces in conventional imaging force microscopy are Pauli repulsion and ionic repulsion. The Pauli exclusion principle forbids the charge clouds of two electrons with the same quantum numbers having any significant overlap; first, the energy of one of the electrons has to be increased. This yields a repulsive force. Additionally, overlap of the charge clouds of electrons can cause an insufficient screening of the nuclear charge, leading to ionic repulsion of coulombic nature. Pauli repulsion and ionic repulsion

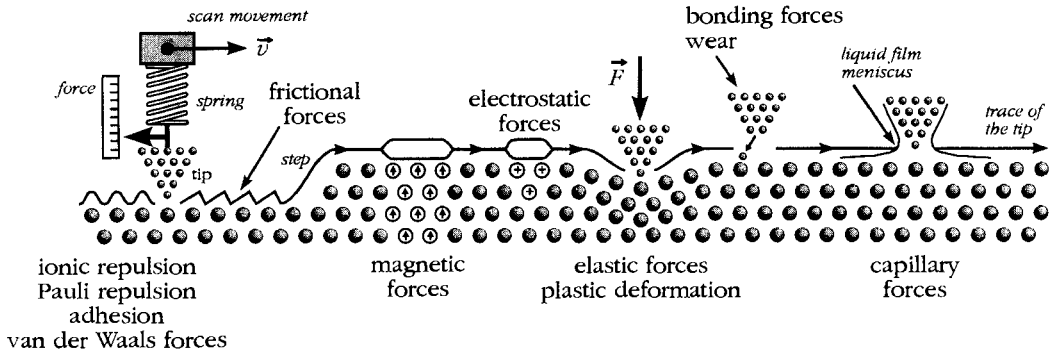


Figure 9. Summary of the forces relevant to SFM.

are very hard, that is, they vary over many orders of magnitude of angstroms [63], thereby preventing the SFM tip from penetrating into the sample. Therefore, only the foremost atoms of the tip interact with the sample surface. This very local interaction enables the imaging of features with atomic or molecular periodicities.

2.3.1.2 Van der Waals Forces

Van der Waals forces are forces between electric dipoles. On the atomic and molecular level, three different types of van der Waals forces are distinguished: (1) the force between two permanent dipoles; (2) the force between a permanent dipole and an induced dipole; and (3) the force between two induced dipoles. The last force arises from fluctuations in the charge distribution of the atoms and molecules, and is known as the dispersion force.

From classical electrodynamics, it is known that dipole–dipole forces are proportional to r^{-7} and, for large distances of more than 10 nm, proportional to $\approx r^{-8}$ due to retardation effects. For assemblies of many dipoles, that is, solid bodies consisting of atoms and molecules which each represent a small dipole, these simple laws change. For example, for a sphere over a flat surface (a geometry which is frequently used to approximate the tip and the sample), the van der Waals force in a vacuum is proportional to r^{-2} if additivity of the forces is assumed [62]. Therefore, van der Waals forces are quite long-range forces which are significant for distances up to some 10 nm (see Fig. 7).

2.3.1.3 Adhesion

All attractive forces occurring in SFM are often termed adhesive forces, including van der Waals forces, capillary forces, bonding forces, etc. Practical models of adhesion, however, do not consider the origin of the forces between surfaces, but describe the nature of the phenomenon in terms of surface energies. In this concept, adhesive forces arise from a reduction in surface energy. If two surfaces are brought together, a certain force has to be applied to separate them. Adhesive forces are mostly of van der Waals origin, except in the case of metals, which can show large adhesive forces due to short-range non-additive electron exchange interactions

giving rise to the formation of metallic bonds [62, 64]. The forces which have to be applied to separate tip and the sample can be derived with the theories of Johnson, Kendall, and Roberts [65] and Derjaguin, Muller, and Toporov [66] for the geometry of a sphere in contact with a flat surface. In air or liquids, surface contaminants and adsorbed molecules reduce the surface energy and, consequently, the occurring adhesive forces [18].

2.3.1.4 Capillary Forces

Under ambient conditions, the dominant attractive forces in force microscopy are capillary forces [19]. Water and other molecules condensing at the sample surface cause the occurrence of an adsorption layer [20, 21]. The SFM tip penetrates through this layer approaching the sample surface. At the tip–sample microcontact, a meniscus is formed which causes an additional attractive force to act on the SFM tip [20]. This force, which depends on the meniscus and hence on the shape of the tip, can be more than 10^{-7} N, reducing the minimal possible loading forces for force microscopy in air to 10^{-9}–10^{-8} N. For soft samples such as biological membranes such forces will have often deformed the sample surface already. However, the loading forces can be reduced to below 10^{-10} N if measured directly in liquids [18]. Alternatively, capillary forces can be avoided by performing the experiments in a glove box with dry gases [19, 20] or in vacuum. In vacuum, however, strong adhesive forces can occur due to the clean surfaces. Attractive forces due to capillarity can be calculated according to the theories of Fogden and White [67] and Maugis and Gauthier-Manuel [68] for the geometry of a sphere over a flat surface.

2.3.1.5 Interatomic and Intermolecular Bonds

Covalent, ionic, or hydrogen bonds can be formed between the tip and the sample, giving rise to an additional attractive force. In extreme cases, for example under ultra-high vacuum conditions on samples showing dangling bonds, this can lead to the destruction of both the tip and the sample [69].

2.3.1.6 Frictional and Elastic Forces

If the tip and the sample are in contact and moved with respect to each other, a frictional force occurs. This force is represented by the component of the force acting on the tip which is parallel to the sample surface, hence causing a torsion of the cantilever which can be measured [32]. Frictional forces can be very large, and in special cases even larger than the normal component of the force. They depend on the surface potential and vary on the atomic scale [49]. Since different materials exhibit different surface potentials, variations of the frictional force can be associated with a different chemical or structural composition of the surface (see Fig. 6) [70].

Not only frictional forces but also elastic forces can provide information about the surface structure. Variations in the local elasticity of the sample have their origin in structural or chemical changes of the investigated sample surface [50]. Experimental set-ups for the measurement

of frictional and elastic forces are described in Sec. 2.2.3.4 of this Chapter.

2.3.1.7 Magnetic and Electrostatic Forces

The magnetic and electrostatic forces are, in comparison with most of the other forces described above, of long-range character and therefore they are most easily measured in noncontact modes. The imaging of magnetic materials with ferromagnetic tips, so-called magnetic force microscopy (MFM), has developed into an important field of force microscopy, and will be described in Sec. 3 of this Chapter. Using similar techniques, trapped charges on insulator surfaces [71] or the domain structure of a ferroelectric material [43, 44] can be visualized.

2.3.2 Contrast Mechanism and Computer Simulations

Although SFM image interpretation seems to be straightforward, many questions concerning the contrast mechanism of SFM and the behavior of atoms close to or at the tip–sample contact area are not yet fully understood. It is not the aim of this chapter to discuss these problems in detail; however, some critical points should be mentioned. A good survey on the theoretical concepts in force microscopy is given by Wiesendanger and Güntherodt [72].

Based on ab initio calculations of the electronic structure, total energy, and forces, Ciraci et al. analyzed the tip–sample interactions of SFM during contact imaging with atomic resolution [52]. They found that at relatively large tip–sample separations, the tip probes the total charge density of the sample surface. However, at small tip–sample separations corresponding to a strong repulsive regime (where most SFM studies are performed), the ion–ion repulsion determines the image contrast. Therefore, the observed maxima in atomically resolved SFM images can be directly attributed to the atomic sites, contrary to STM, where such an interpretation is not generally applicable.

Tekman and Ciraci [74] showed that even in SFM measurements performed with blunt tips, features with the periodicity of the atomic lattice can be resolved although several atoms are involved in the process of contrast formation, thus preventing the observation of point-like defects. Energetic considerations suggest that a single atom at the tip end is not stable at loading forces which are practicable under ambient conditions or in an ultra-high vacuum ($\geqslant 10^{-9}$ N) [74]. This might explain why scanning force micrographs of atomically resolved surfaces usually show only defect-free atomic lattice structures, in contrast to scanning tunneling micrographs, on which defects and kinks can frequently be observed. Ohnesorge and Binnig [75], measuring in liquids, demonstrated by resolving atomic scale kinks that attractive forces of about 10^{-11} N can be used to obtain true atomic resolution. Nevertheless, since other authors have reported the observation of defects in standard SFM contact imaging [76], this problem has not yet been satisfactorily resolved.

Apart from the theoretical analysis, computer simulations can provide insights into the physics of the tip–sample system on the atomic level. The evolution of such

a system can be simulated with high temporal and spatial resolution via a direct numerical solution of the model equations of motion, employing a realistic interatomic interaction potential. Examples of such studies are given in the literature [77].

Acknowledgments

It is a pleasure to thank W. Allers, H. Bluhm, M. Seider, and O. Zwörner for supplying SFM images for this chapter, and H. Eggers and P. Köster for electron micrographs. Moreover, I would like to acknowledge the people at Hamburg University who provided assistance and technical support.

2.4 References

[1] G. Binnig, H. Rohrer, C. Gerber, E. Weibel, *Phys. Rev. Lett.* **1982,** *49*, 57; G. Binnig, H. Rohrer, *Helv. Phys. Acta* **1982,** *55*, 726.

[2] H. K. Wickramasinghe, *J. Vac. Sci. Technol. A* **1989,** *8*, 363.

[3] G. Binnig, C. F. Quate, C. Gerber, *Phys. Rev. Lett.* **1986,** *56*, 930.

[4] G. Meyer, N. M. Amer, *Appl. Phys. Lett.* **1988,** *53*, 1045.

[5] S. Alexander, L. Hellemans, O. Marti, J. Schneir, V. Elings, P. K. Hansma, M. Longmire, J. Gurley, *J. Appl. Phys.* **1989,** *64*, 164.

[6] D. Rugar, H. J. Mamin, P. Guethner, *Appl. Phys. Lett.* **1989,** *55*, 2588.

[7] E. C. Teague, F. E. Scire, S. M. Baker, S. W. Jensen, *Wear* **1982,** *83*, 1.

[8] U. D. Schwarz, H. Haefke, T. Jung, E. Meyer, H.-J. Güntherodt, R. Steiger, J. Bohonek, *Ultramicroscopy* **1992,** *41*, 435; U. D. Schwarz, H. Haefke, H.-J. Güntherodt, J. Bohonek, R. Steiger, *J. Imaging Sci. Technol.* **1993,** *37*, 344.

[9] E. Henderson, P. G. Haydon, D. S. Sakaguchi, *Science* **1992,** *257*, 1944; J. H. Hoh, P. K. Hansma, *Trends Cell Biol.* **1992,** *2*, 208; J. H. Hoh, G. E. Sosinsky, J.-P. Revel, P. K. Hansma, *Biophys. J.* **1993,** *65*, 149.

[10] D. Anselmetti, R. Lüthi, E, Meyer, T. Richmond, M. Dreier, J. E. Frommer, H.-J. Güntherodt, *Nanotechnology* **1994,** *5*, 87.

[11] G. Friedbacher, P. K. Hansma, E. Ramli, G. D. Stucky, *Science* **1991,** *253*, 1261.

[12] U. D. Schwarz, H.Haefke, *Appl. Phys. A* **1994,** *59*, 33.

[13] R. Wigren, H. Elwing, R. Erlandsson, S. Welin, I. Lundström, *FEBS Lett.* **1991,** *280*, 225; W. A. Rees, R. W. Keller, J. P. Vesenka, G. Yang, C. Bustamante, *Science* **1993,** *260*, 1646.

[14] J. Frommer, *Angew. Chem., Int. Ed. Engl.* **1992,** *31*, 1298; J. Frommer, R. Lüthi, E. Meyer, D. Anselmetti, M. Dreier, R. Overney, H.-J. Güntherodt, M. Fujihira, *Nature* **1993,** *364*, 198.

[15] S. Karrasch, M. Dolder, F. Schabert, J. Ramsden, A. Engel, *Biophys. J.* **1993,** *65*, 2437.

[16] M. Hipp, H. Bielefeld, J. Colchero, O. Marti, J. Mlynek, *Ultramicroscopy* **1992,** *42–44*, 1498.

[17] B. Drake, C. B. Prater, A. L. Weisenhorn, S. A. C. Gould, T. R. Albrecht, C. F. Quate, D. S. Cannell, H. G. Hansma, P. K. Hansma, *Science* **1989,** *243*, 1586.

[18] A. L. Weisenhorn, P. Maivald, H.-J. Butt, P. K. Hansma, *Phys. Rev. B* **1992,** *45*, 11226.

[19] T. Thundat, X.-Y. Zheng, G. Y. Chen, R. J. Warmack, *Surf. Sci. Lett.* **1993,** *294*, L939.

[20] M. Binggeli, C. M. Mate, *Appl. Phys. Lett.* **1994,** *65*, 415.

[21] J. Hu, X.-D. Xiao, D. F. Ogletree, M. Salmeron, *Science* **1995,** *268*, 267.

[22] G. Meyer, N. M. Amer, *Appl. Phys. Lett.* **1990,** *56*, 2100; L. Howald, H. Haefke, R. Lüthi, E. Meyer, G. Gerth, H. Rudin, H.-J. Güntherodt, *Phys. Rev. B* **1994,** *49*, 5651.

[23] G. Neubauer, S. R. Cohen, G. M. McClelland, D. Horne, C. M. Mate, *Rev. Sci. Instrum.* **1990,** *61*, 2296.

[24] F. J. Giessibl, *Jpn. J. Appl. Phys.* **1994,** *33*, 3726.

[25] F. J. Giessibl, G. Binnig, *Ultramicroscopy* **1992,** *42–44*, 281; T. R. Albrecht, P. Grütter, D. Rugar, *Ultramicroscopy* **1992,** *42–44*, 1638.

[26] H. Heinzelmann, E. Meyer, H. Rudin, H.-J. Güntherodt in *Scanning Tunneling Microscopy and Related Methods* (Eds.: R. J. Behm, N. Garcia, H. Rohrer), Kluwer, Dordrecht **1990,** p. 443; D. Rugar, P. Hansma, *Phys. Today* **1990,** *Oct.*, 23; D. Sarid, V. Elings, *J. Vac. Sci. Technol. B* **1991,** *9*, 431; J. Frommer, E. Meyer, *J. Phys. C: Condens. Matter* **1991,** *3*, S1; C. F. Quate, *Surf. Sci.* **1994,** *299/300*, 980.

[27] D. Sarid, *Scanning Force Microscopy*, 1st Ed., Oxford University Press, New York **1991**; R. Wiesendanger, H.-J. Güntherodt (Eds.), *Scanning Tunneling Microscopy II, Springer Series in Surface Sciences*, Vol. 28, 1st Ed., Springer-Verlag, Berlin **1992**; R. Wiesendanger, H.-J. Güntherodt (Eds.), *Scanning Tunneling Microscopy III, Springer Series in Surface Sciences*,

Vol. 29, 1st Ed., Springer-Verlag, Berlin **1993**; R. Wiesendanger, *Scanning Probe Microscopy and Spectroscopy*, Cambridge University Press, Cambridge **1994**.

[28] D. W. Pohl, *IBM J. Res. Develop.* **1986,** *30*, 417; S. Park, C. F. Quate, *Rev. Sci. Instrum.* **1987,** *58*, 2004; Y. Kuk, P. J. Silverman, *Rev. Sci. Instrum.* **1989,** *60*, 165.

[29] T. R. Albrecht, C. F. Quate, *J. Vac. Sci. Technol. A* **1988,** *6*, 271.

[30] T. R. Albrecht, S. Akamine, T. E. Carver, C. F. Quate, *J. Vac. Sci. Technol. A* **1990,** *8*, 3386; O. Wolter, T. Bayer, J. Greschner, *J. Vac. Sci. Technol. B* **1991,** *9*, 1353.

[31] U. D. Schwarz, H. Haefke, P. Reimann, H.-J. Güntherodt, *J. Microsc.* **1994,** *173*, 183.

[32] O. Marti, J. Colchero, J. Mlynek, *Nanotechnology* **1990,** *1*, 141; G. Meyer, N. M. Amer, *Appl. Phys. Lett.* **1990,** *57*, 2089.

[33] Y. Martin, C. C. Williams, H. K. Wickramasinghe, *J. Appl. Phys.* **1987,** *61*, 4723.

[34] G. M. McClelland, R. Erlandsson, S. Chiang in *Review of Progress in Quantitative Non-Destructive Evaluation*, Vol. 6 (Eds.: D. O. Thompson, D. E. Chimenti), Plenum Press, New York **1987,** p. 1307; A. J. den Boef, *Appl. Phys. Lett.* **1989,** *55*, 439; C. Schönenberger, S. F. Alvarado, *Rev. Sci. Instrum.* **1989,** *60*, 3131.

[35] P. J. Mulhern, T. Hubbard, C. S. Arnold, B. L. Blackford, M. H. Jericho, *Rev. Sci. Instrum.* **1991,** *62*, 1280; A. Moser, H. J. Hug, T. Jung, U. D. Schwarz, H.-J. Güntherodt, *Measure. Sci. Technol.* **1993,** *4*, 769.

[36] C. A. J. Putnam, B. G. de Grooth, N. F. van Hulst, J. Greve, *Ultramicroscopy* **1992,** *42–44*, 1509.

[37] D. Sarid, D. Iams, V. Weissenberger, L. S. Bell, *Opt. Lett.* **1988,** *13*, 1057.

[38] D. Sarid, D. A. Iams, J. T. Ingle, V. Weissenberger, J. Ploetz, *J. Vac. Sci. Technol. A* **1990,** *8*, 378.

[39] M. Tortonese, R. C. Barrett, C. F. Quate, *Appl. Phys. Lett.* **1993,** *62*, 834.

[40] F. J. Giessibl, *Science* **1995,** *267*, 68.

[41] T. R. Albrecht, C. F. Quate, *J. Appl. Phys.* **1987,** *62*, 2599; H. Heinzelmann, P. Grütter, E. Meyer, H.-R. Hidber, L. Rosenthaler, M. Ringger, H.-J. Güntherodt, *Surf. Sci.* **1987,** *189/190*, 29; O. Probst, S. Grafström, J. Kowalski, R. Neumann, M. Wörtge, *J. Vac. Sci. Technol. B* **1991,** *9*, 626.

[42] T. Göddenhenrich, H. Lemke, U. Hartmann, C. Heiden, *J. Vac. Sci. Technol. A* **1990,** *8*, 383.

[43] R. Lüthi, E. Meyer, L. Howald, H. Haefke, D. Anselmetti, M. Dreier, M. Rüetschi, T. Bonner, R. M. Overney, J. Frommer, H.-J. Güntherodt, *J. Vac. Sci. Technol. B* **1994,** *12*, 1673.

[44] F. Saurenbach, B. D. Terris, *Appl. Phys. Lett.* **1990,** *56*, 1703.

[45] T. R. Albrecht, P. Grütter, D. Horne, R. Rugar, *J. Appl. Phys.* **1991,** *69*, 668.

[46] U. Hartmann in *Scanning Tunneling Microscopy III* (Eds.: R. Wiesendanger, H.-J. Güntherodt), *Springer Series in Surface Sciences*, Vol. 29, 1st Edn., Springer-Verlag, Berlin **1993,** p. 293.

[47] R. Overney, E. Meyer, *MRS Bull.* **1993,** *19*, 26.

[48] S. Grafström, M. Neitzert, T. Hagen, J. Ackermann, R. Neumann, O. Probst, M. Wörtge, *Nanotechnology* **1993,** *4*, 143.

[49] C. M. Mate, G. M. McClelland, R. Erlandsson, S. Chiang, *Phys. Rev. Lett.* **1987,** *59*, 1942; R. Erlandsson, G. Hadziioannou, C. M. Mate, G. M. McClelland, S. Chiang, *J. Chem. Phys.* **1988,** *89*, 5190; R. M. Overney, H. Takano, M. Fujihira, *Phys. Rev. Lett.* **1994,** *72*, 3546; S. Fujisawa, E. Kishi, Y. Sugawara, S. Morita, *Phys. Rev. B* **1995,** *51*, 7849.

[50] M. Rademacher, R. W. Tillmann, M. Fritz, H. E. Gaub, *Science* **1992,** *257*, 1900.

[51] Q. Zhong, D. Inniss, K. Kjoller, V. B. Elings, *Surf. Sci. Lett.* **1993,** *290*, L688.

[52] S. Ciraci, A. Baratoff, I. P. Batra, *Phys. Rev. B* **1990,** *41*, 2763.

[53] F. O. Goodman, N. Garcia, *Phys. Rev. B* **1991,** *43*, 4728.

[54] A. Torii, M. Sasaki, K. Hane, S. Okuma, *Sensors Actuators A* **1994,** *44*, 153.

[55] H. J. Hug, A. Moser, T. Jung, O. Fritz, A. Wadas, I. Parashnikov, H.-J. Güntherodt, *Rev. Sci. Instrum.* **1993,** *64*, 2920; H. J. Hug, A. Moser, O. Fritz, I. Parashnikov, H.-J. Güntherodt, T. Wolf, *Physica B* **1994,** *194–196*, 377.

[56] U. Dammer, O. Popescu, P. Wagner, D. Anselmetti, H.-J. Güntherodt, G. N. Misevic, *Science* **1995,** *267*, 1173.

[57] C. D. Frisbie, L. F. Rozsnyai, A. Noy, M. S. Wrighton, C. M. Lieber, *Science* **1994,** *265*, 2071.

[58] H. K. Christenson, *J. Coll. Interface Sci.* **1988,** *121*, 170.

[59] T. A. Jung, A. Moser, H. J. Hug, D. Brodbeck, R. Hofer, H. R. Hidber, U. D. Schwarz, *Ultramicroscopy* **1992,** *42–44*, 1446; Y. Kim, C. M. Lieber, *Science* **1992,** *257*, 375; G. Persch, C. Born, B. Utesch, *Appl. Phys. A* **1994,** *59*, 29; B. Bhushan, J. N. Israelachvili, U. Landmann, *Nature* **1995,** *374*, 607.

[60] T. A. Jung, A. Moser, M. T. Gale, H. J. Hug, U. D. Schwarz in *Technology of Proximal Probe Lithography*, Vol. IS 10 (Ed.: C. R. K. Marrian), SPIE Institutes for Advanced Optical Technologies, Bellingham **1993,** p. 234.

[61] A. Majumdar, P. I. Oden, J. P. Carrejo, L. A. Nagahara, J. J. Graham, J. Alexander, *Appl. Phys. Lett.* **1992,** *61*, 2293; E. S. Snow, P. M. Campell, *Appl. Phys. Lett.* **1994,** *64*, 1932; L. Tsau, D. Wang, K. L. Wang, *Appl. Phys. Lett.* **1994,** *64*, 2133; M. Wendel, S. Kühn, H. Lorenz, J. P. Kotthaus, M. Holland, *Appl. Phys. Lett.* **1994,** *65*, 1775.

[62] J. N. Israelachvili, *Intermolecular and Surface Forces*, 2nd Edn., Academic Press, London **1991**.

[63] R. G. Gordon, Y. S. Kim, *J. Chem. Phys.* **1972,** *56*, 3122.

[64] U. Dürig and A. Stalder, in *Physics of Sliding Friction* (Eds.: B. N. J. Persson and E. Tosatti), NATO ASI Series, Serie E: Applied Sciences, Vol. 311, Kluwer Academic Publishers, Dordrecht, **1996,** p. 266.

[65] K. L. Johnson, K. Kendall, A. D. Roberts, *Proc. R. Soc. Lond. A* **1971,** *324*, 301.

[66] B. V. Derjaguin, V. M. Muller, Y. P. Toporov, *J. Colloid Interface Sci.* **1975,** *53*, 314; B. D. Hughes, L. R. White, *Q. J. Mech. Appl. Math.* **1979,** *33*, 445; V. M. Muller, V. S. Yushenko, B. V. Derjaguin, *J. Colloid Interface Sci.* **1980,** *77*, 91; **1983,** *92*, 92; D. Maugis, *J. Colloid Interface Sci.* **1992,** *150*, 243.

[67] A. Fogden, L. R. White, *J. Colloid Interface Sci.* **1990,** *138*, 414.

[68] D. Maugis, B. Gauthier-Manuel, *J. Adhesion Sci. Technol.* **1994,** *8*, 1311.

[69] L. Howald, R. Lüthi, E. Meyer, P. Güthner, H.-J. Güntherodt, *Z. Phys. B* **1994,** *93*, 267.

[70] R. M. Overney, E. Meyer, J. Frommer, D. Brodbeck, R. Lüthi, L. Howald, H.-J. Güntherodt, M. Fujihira, H. Takano, Y. Gotoh, *Nature* **1992,** *359*, 133.

[71] C. Schönenberger, S. F. Alvarado, *Phys. Rev. Lett.* **1990,** *65*, 3162.

[72] R. Wiesendanger, H.-J. Güntherodt (Eds.), *Scanning Tunneling Microscopy III, Springer Series in Surface Sciences*, Vol. 29, 1st Edn., Springer-Verlag, Berlin **1993**.

[73] E. Tekman, S. Ciraci, *J. Phys. C: Condens. Matter* **1991,** *3*, 2613.

[74] C. F. Quate, *Surf. Sci.* **1994,** *299/300*, 980.

[75] F. Ohnesorge, G. Binnig, *Science* **1993,** *260*, 1451.

[76] S. S. Sheiko, M. Möller, E. M. C. M. Reuvekamp, H. W. Zandbergen, *Phys. Rev. B* **1993,** *48*, 5675.

[77] U. Landman, W. D. Luedtke, M. W. Ribarsky, *J. Vac. Sci. Technol. A* **1989,** *7*, 2829; U. Landman, W. D. Luedtke, A. Nitzan, *Surf. Sci.* **1989,** *210*, L177; U. Landman, W. D. Luedtke, N. A. Burnham, R. J. Colton, *Science* **1990,** *248*, 454; J. Belak, I. F. Stowers in *Fundamental of Friction: Macroscopic and Microscopic Processes* (Eds.: I. L. Singer, H. M. Pollock), Kluwer, Dordrecht **1992,** p. 511; J. A. Harrison, C. T. White, R. J. Colton, D. W. Brenner, *Surf. Sci.* **1992,** *271*, 57; *Phys. Rev. B* **1992,** *46*, 9700; *J. Phys. Chem.* **1993,** *97*, 6573; U. Landman, W. D. Luedtke, *MRS Bull.* **1993,** *18*, 993.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

3 Magnetic Force Microscopy

3.1 Introduction

Magnetic force microscopy (MFM) is the third generation of scanning probe techniques after scanning tunneling microscopy (STM) and scanning force microscopy (SFM) [1]. MFM has been designed to study magnetic materials on the nanometer scale. The first results obtained by this method were reported by Martin and Wickramasinghe in 1987 [2]. The technique relies on a magnetostatic interaction between a magnetic sample and a probing sensor. Figure 1 is an illustration of MFM operation. The magnetic tip, which acts as the sensor, is mounted on a flexible cantilever. It is raster scanned over the sample surface, typically in the range from 20 to hundreds of nanometers. A magnetic sample with a domain structure produces a complicated stray field over the surface. The aim of MFM is to map the stray field as close to the surface as possible. The interaction which occurs when a sample is scanned by an MFM tip is measured via a deflection of the cantilever. There are many modes of MFM operation and various techniques to monitor the bending of the cantilever. There are, however, only two physical quantities directly measured by MFM: the force or force gradient acting on a magnetic tip.

3.2 Force Measurement

Figure 1 is a sketch of the set-up of one particular method used to measure magnetic forces. The method of detecting cantilever bending shown in Fig. 1 is called a deflection detection scheme, and was invented by Meyer and Amer in 1988 [3]. A collimated laser beam is focused onto the back face of the cantilever and is reflected toward a set of photodetectors. In the simplest case there are two photodetectors which monitor the force causing an upward or downward cantilever deflection. One photodetector collects more light than the other one due to the deflection of the cantilever. Photocurrents produced by the photodetectors are fed to a differential amplifier. Finally, a signal is obtained that is proportional to the difference in light detected by the photodetectors, and which is a measure of the cantilever deflection. Usually, instead of two photodetectors there is a set of four forming a quadrant as in Fig. 1. This allows additional measurement of lateral forces due to the torque applied to the cantilever. However, we will concentrate on normal forces, which have been the main interest of MFM groups. Since we know now how to detect forces, the next step is to perform MFM. This means that a magnetic tip has to move

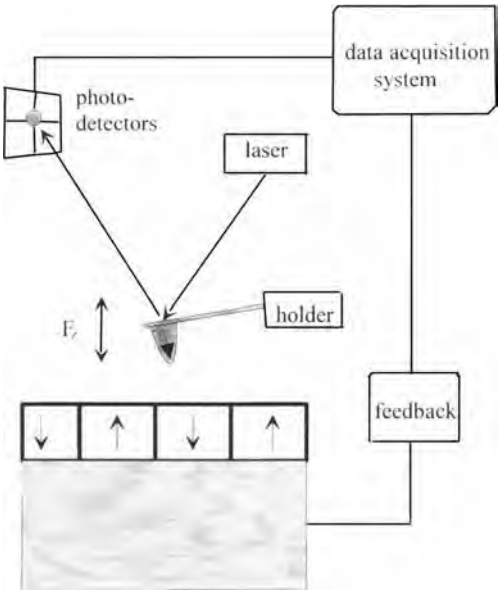


Figure 1. A sketch of the experimental set-up used to perform MFM.

over a sample surface at a well-defined height despite any topographic obstacles. This has been usually called the constant height mode. A very reliable way to do this has been introduced by Digital Instruments (Santa Barbara, California, U.S.A.) in one of their commercial microscopes. Figure 2 outlines the principle of their

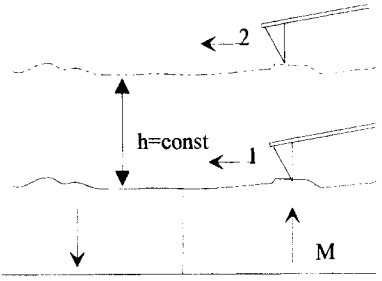


Figure 2. An illustration of the constant height mode as introduced by Digital Instruments.

method, called the 'lift mode'. First, the tip is brought into contact with the surface of the sample and performs one scan line. After the topography of this scan line is stored, the tip is scanned along a path parallel to the previously stored scan line at a predefined height from the sample surface. During this second scan the tip–sample separation is nominally constant. Deflections of the cantilever due to magnetic forces are then plotted as a function of position. Such a procedure provides both the topography and the magnetic force images of the same area. This type of operation has been used to study either soft magnetic materials such as garnet films and permalloy or hard magnetic materials such as barium ferrite or hard disk [4, 5]. Figure 3 is an example of an

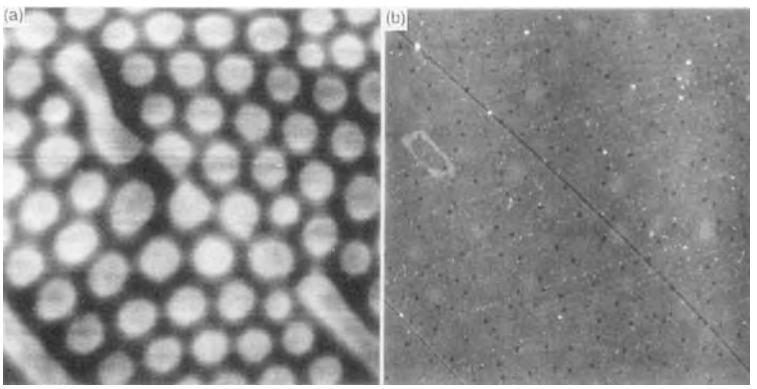


Figure 3. The MFM image of (a) bubble domains in a garnet and (b) the corresponding topographic image.

MFM image with a corresponding topographic scan obtained by the lift mode technique on an epitaxially grown $(YSmLuCa)_3(FeGe)_5O_{12}$ garnet film with a bubble structure [4]. Bright circles in the MFM image represent cylindrical domains with a magnetic polarization directed perpendicularly to the surface and opposite to a surrounding area. The bright areas indicate that the MFM tip is being repelled by the sample.

The general formula describing the interaction between a tip and a magnetic sample can be written as

$$E = -\iiint_{\text{tip}} H(r)M \, dV \tag{1}$$

where M is the tip magnetization and $H(r)$ is the magnetic field above a sample. A force acting on the tip along the z-axis (perpendicular to the surface) is related to the interaction energy by

$$F_z = -\frac{\partial E}{\partial z} \tag{2}$$

The force F_z is related directly to the bending of the cantilever detected during MFM operation by Hooke's law:

$$F_z = -kz \tag{3}$$

where k denotes the spring constant of a cantilever.

A detailed calculation of forces which subsequently can be compared with experimental data has been done with a well-defined tip geometry and distribution of the stray field $H(r)$. Different shapes for the top were considered: conical, pyramidal, and cylindrical. These theoretical shapes have their experimental counterparts. Examples of the calculations of forces or force gradients giving an analytical formula or a numerical approach for a well-defined domain structure can be found in some references [6–11]. We have touched here on one of the fundamental problems of MFM, that is, to describe the tip shape, which is in practice determined by the tip preparation. An MFM tip has to fulfill certain conditions. It should be as sensitive as possible, nonevasive, and easily available. Since in magnetism we are dealing with a variety of materials with different properties it is very difficult, if not impossible, to prepare such a universal tip which will give the best MFM image. However, a search was undertaken, completed with some success to find such a tip [5]. It should be mentioned that each particularly soft magnetic sample needs a series of experiments to find the most appropriate MFM tip.

Historically, the first tips were made from magnetic wires of, for example, iron or nickel [2, 11, 12]. The end of the wire was chemically etched to form a sharp needle. Some tens or hundreds of micrometers at the very end of the wire was bent toward a sample. Such a piece of wire with an 'L' shape then acted as a cantilever with the tip at one end. Even if these tips were not magnetized they produced quite a strong signal. Due to their shape anisotropy (a long cylinder with a cone at the end), a large part of the volume probably formed a single magnetic domain responsible for the effective interaction with a magnetic sample. The word 'probably' is used here to highlight an important fact. It was not possible to determine exactly the status of the magnetic domain structure within an MFM tip. The best and most practical way to describe a newly prepared tip is to use it on a well-defined domain structure. A piece of a standard hard disk taken from a computer is a good test

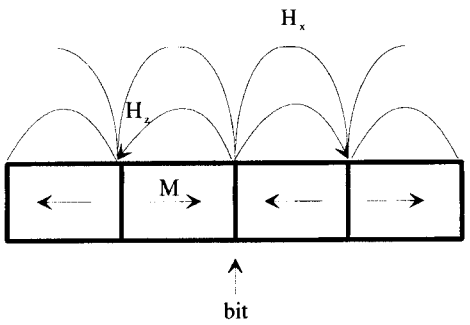


Figure 4. The magnetic structure of a hard disk, used frequently in MFM as a test sample.

sample. Its structure is shown in Fig. 4. The stray field emerging from a hard disk has a strong in-plane component H_x and a strong component H_z just above a bit (transition area). The MFM image obtained of a hard disk gives us information about the effective x- and z-components of the tip magnetization since the energy of interaction described in Eq. (1) directly relates a measured force to the components of tip magnetization.

Magnetic wires used as cantilevers with tips have two important disadvantages. The first one is that the spring constant of the cantilever is not well defined. It always differs from preparation to preparation. In addition, etching the end might form a different domain structure inside the wire. This, however, can be partially overcome by magnetizing the tip, which then leads to the second important disadvantage of wire tips, that is, a strong stray field emerging from such tips [13]. In order to overcome this difficulty, it is necessary to scan over a sample at a sufficiently large distance, which, however, leads to a reduction in lateral resolution. A very logical and natural way to solve these problems is to completely abandon wire tips. The other choice is to use a magnetic thin film deposited on a nonmagnetic cantilever. Experimental and theoretical results for thin film tips came almost at the same time [14, 15].

Thin film tips were first made by sputtering iron or CoZrNb on tungsten wires, giving very good results on PtCo multilayers [14]. Magneto-optically written domains in these multilayers have been imaged. Next, thin film tips were prepared on microfabricated silicon cantilevers with integrated tips [16]. This procedure assured a well-defined spring constant and a well-defined geometry for the nonmagnetic tip. Up to now there have been different magnetic materials used for the coatings, such as iron, cobalt, permalloy, nickel, and CoPtCr alloys, deposited either by sputtering or evaporation.

The newest state-of-the-art tips are also magnetic thin film tips, but deposited on so-called contamination tips or EBD (electron beam-deposited) tips [17]. It was found that an electron beam of a scanning electron microscope focused on a substrate grows pillars containing mainly carbon. These pillars have the shape of cones with a radius of curvature at the apex between 20 and 40 nm and a cone half-angle of 3–50° [17]. MFM tips prepared by the evaporation of iron on such small cones give a smaller MFM signal than other tips but always have a very good lateral resolution of around 50 nm. They have been used to study garnet films never overwriting a domain structure. The usual thickness of the iron layer has been in the range from 10 to 20 nm. The best MFM results have been obtained with the double-layer thin film tips described earlier [5].

3.3 Force Gradient Measurement

As we mentioned above, the other physical quantity measured in MFM is the gradient of a force. This is realized in the so-called alternating current (AC) mode. If we consider the set-up in Fig. 1, some important changes have been made to perform the AC mode. First, there is a bimorph mounted instead of the part labeled 'holder' in Fig. 1. The bimorph is used to oscillate the cantilever near its resonance frequency. As the vibrating cantilever approaches the sample surface, the tip is influenced by an overall force gradient. This means that parallel to the gradient of magnetic forces the tip encounters a gradient of attractive van der Waals forces. Since a force applied to a cantilever is proportional to its bending [Eq. (3)] the force gradient is going to change the effective spring constant, k_{eff}, of the cantilever. If we consider the harmonic oscillation of a cantilever, then the resonance frequency ω_0 is given by the formula

$$\omega_0 = \sqrt{\frac{k_{\text{eff}}}{m}} \tag{4}$$

where m is the effective mass of the cantilever. The effective spring constant is $k_{\text{eff}} = k - F'$, where F' is the overall force gradient acting on the cantilever and k is the nominal spring constant of the cantilever.

The change in the force gradient, expected when the tip is approaching the sample surface, shifts the resonance frequency, which subsequently alters the amplitude of the spring vibration (Fig. 5). This can be detected using a laser heterodyne probe [2, 12]. If the tip is then

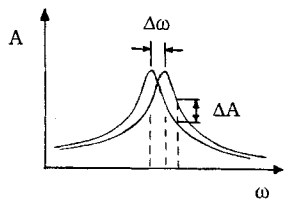


Figure 5. The shift in the resonance frequency, $\Delta\omega$, caused by an external force gradient. A change in frequency results in a change of the amplitude, ΔA, at a given frequency.

raster scanned over the surface at a constant height, as described earlier, we detect a force gradient in the constant height mode. This method of operation was introduced by Martin and Wickramasinghe in the first MFM report [2]. The sample they investigated was a thin film magnetic recording head (Fig. 6). The current of the head was modulated at a certain frequency, producing a modulated stray field of the same frequency. The vibration of the cantilever due to the modulated head field was detected using a heterodyne interferometer which was capable of detecting a vibration amplitude down to 5×10^{-5} Å in a bandwidth of 1 Hz. The detected amplitude signal from the laser heterodyne probe is then proportional to the magnetic force gradient over the recording head.

The other method of MFM operation is the constant force gradient mode, in which the signal from the laser probe is compared with an electronic reference. A feedback circuit tries to adjust the tip–sample separation by adjusting the z-signal of a piezo element upon which a sample is mounted (Fig. 1) to maintain a constant force gradient. Figure 7 shows an MFM image obtained in this way by van Kesteren et al. [12]. The image reveals an array of thermomagnetically written cylindrical

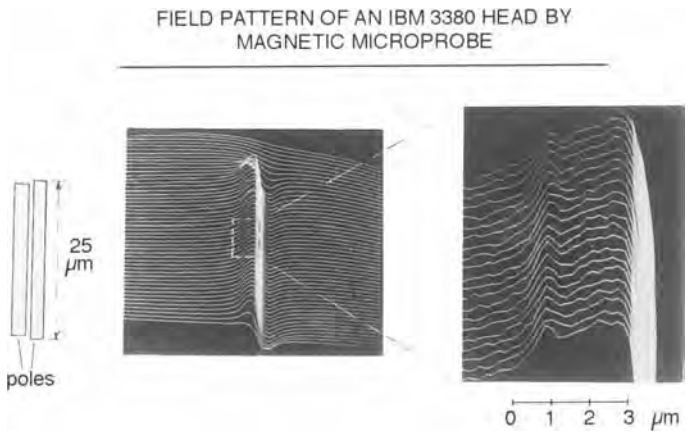


Figure 6. Magnetic force gradient contours of a thin film recording head excited by AC as measured by MFM. (Courtesy of Martin and Wickramasinghe [2].)

domains in Co/Pt multilayer films. Both of the examples in Figs. 6 and 7 show the direct use of MFM in magnetic recording technology.

The same technique based on the detection of a change in the oscillation amplitude, ΔA, has been used to study permalloy films [18]. Again, the constant force gradient was traced in order to get an MFM image. In this case, however, the authors used a fiber-optic interferometer to sense the deflection of the cantilever. One of the advantages of this system over a heterodyne interferometer is its lower thermal drift, which usually affects the optical path length of a laser beam.

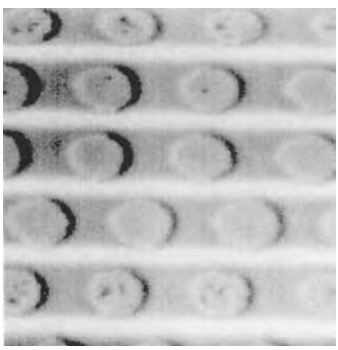


Figure 7. MFM images of a laser-written array of domains in a Co/Pt multilayer. (Courtesy of van Kesteren et al. [12].)

The fiber which carries the light is placed some tens of micrometers above the cantilever. A small part of the light is reflected by the end of the fiber. This light serves as a reference signal. The rest of the light passes the end of the fiber and spreads toward the cantilever. Again, only a part of the light reflected from the back of the cantilever enters the fiber. This fraction of the light is partly determined by the fiber–cantilever spacing. The interference of both reflected beams, by the end of the fiber and by the cantilever, determines the interferometer response [19].

Figure 8 shows the experimental data obtained by Mamin et al. on 2.3 μm thick permalloy film [18]. The sample was placed on an electrode with an applied voltage of 2–10 V, providing an offset of the electrostatic force gradient. This ensures the presence of a net attractive interaction, giving stable experimental conditions. Magnetic domains (bright areas) are surrounded by dark or bright lines representing domain walls. The arrow indicates a change in the domain wall contrast. The

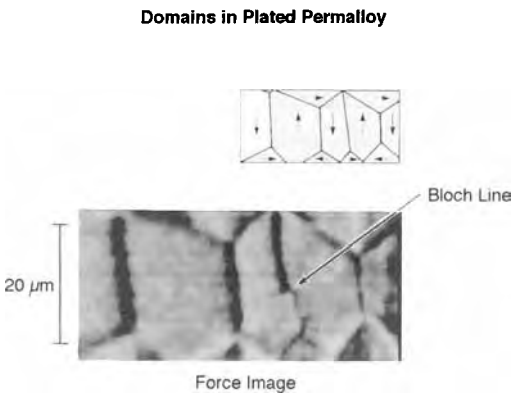


Figure 8. MFM image of domains in a permalloy film. The arrow indicates the location of a Bloch line. (Courtesy of Mamin et al. [18].)

authors correlate this change with the position of a Bloch line.

Experiments on both TbFe and permalloy have been performed using an electrochemically etched iron wire cantilever. The experiment on permalloy showed that MFM is able to recognize the position of a domain wall and, additionally, is able to detect the sense of rotation of the magnetization across the domain wall. This demonstrates that MFM is an important tool to study the internal structure of domain walls.

Another example using the AC technique is shown in Fig. 9 [20]. The authors measured a change in the oscillation amplitude of a vibrating cantilever to study a Co–Pd multilayer. They mapped constant force gradient contours. Figure 9 shows static images of a dynamic process which occurs when the magnetic tip is scanned over the sample at a close tip–sample distance. The image Fig. 9a, taken at a higher distance than in Fig. 9b, shows domain contrast, whereas the image in Fig. 9b, observed closer to the surface, presents a domain wall contrast of the same area. Domain walls are visible as dark lines.

All of the presented examples show various means of studying different aspects of magnetism using MFM based on the optical detection of a cantilever deflection. However, there are also other ways to

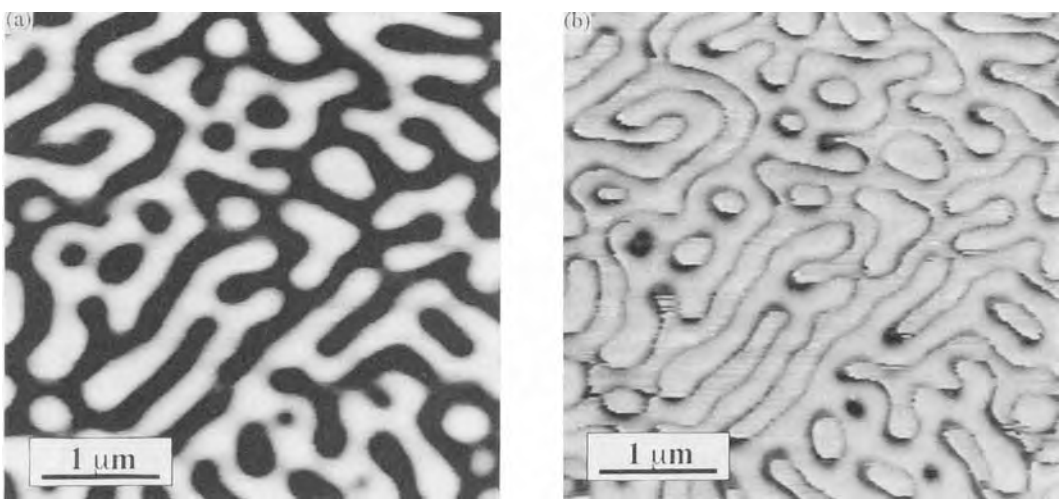


Figure 9. MFM images of the domain structure of a Co–Pd multilayer measured (a) at 90 nm and (b) at 40 nm tip–sample distances, respectively. (Courtesy of Barnes et al. [20].)

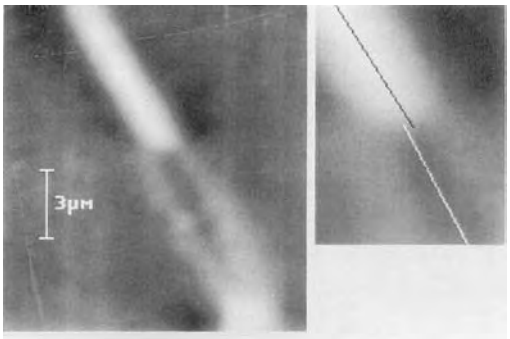


Figure 10. MFM contrast image of a 180° Bloch wall in an iron whisker. (Courtesy of Göddenhenrich et al. [21].)

monitor cantilever movement. Göddenhenrich et al. developed a microscope with a capacitively controlled cantilever displacement [21]. A small plate of thin aluminum foil attached to the back of a lever served as a capacitor plate. The opposite plate was mounted on a tubular control piezo element. The deflection of the cantilever due to the forces acting on the magnetic tip caused a change in the capacitance. The sensitivity of such a system is 10^{-6} pF, which corresponds to a change of 0.01 Å in the distance between the two plates. The force detection limit is of the order of 1×10^{-10} N [21]. Figure 10 shows an MFM image obtained over a domain wall in a single-crystal iron whisker using the capacitive detection scheme. The change in the domain wall contrast was observed (dark to bright) due to the different polarizations of the magnetization within the domain wall.

The other nonoptical way to detect cantilever deflection is well known under the name of tunneling stabilized magnetic force microscopy (TSMFM). TSMFM is performed by using a scanning tunneling microscope with a flexible, magnetic tunneling tip instead of the usual rigid tip. The tip position is stabilized near the surface of the magnetic sample using the STM feedback system as tunneling occurs between the tip and sample surface. If the stray field from the sample attracts the tip, then the feedback system maintaining a constant tunneling gap pulls the tip off the surface. Numerous examples using this method can be found in the references [22–24].

All of the images presented in this chapter provide evidence of the potential of MFM either to study fundamental problems in magnetism or for direct use in technology.

Acknowledgments

I would like to thank J. R. Barnes, U. Hartmann, H. W. van Kesteren, H. J. Mamin, and Y. Martin, who provided illustrations from their work for this article.

3.4 References

[1] G. Binnig, H. Rohrer, *Helv. Phys. Acta* **1982,** *55*, 726.
[2] Y. Martin, H. K. Wickramasinghe, *Appl. Phys. Lett.* **1987,** *50*, 1455.
[3] G. Meyer, N. M. Amer, *Appl. Phys. Lett.* **1988,** *53*, 2400; N. M. Amer, G. Meyer, *Bull. Am. Phys. Soc.* **1988,** *33*, 319.
[4] A. Wadas, J. Moreland, P. Rice, R. R. Katti, *Appl. Phys. Lett.* **1994,** *64*, 1156.
[5] A. Wadas. P. Rice, J. Moreland, *Appl. Phys. A* **1994,** *59*, 63.
[6] A. Wadas, P. Grütter, *Phys. Rev. B* **1989,** *39*, 1013.
[7] U. Hartmann, *Phys. Lett.* **1989,** *137*, 475.
[8] U. Hartmann, *Phys. Rev. B* **1989,** *40*, 7421.
[9] U. Hartmann, *J. Vac. Sci. Technol. A* **1990,** *8*, 411.
[10] A. Wadas, P. Grütter, H.-J. Güntherodt, *J. Appl. Phys.* **1990,** *67*, 3462.

[11] D. Rugar, H. J. Mamin, P. Güthner, S. E. Lambert, J. E. Stern, I. McFadyen, T. Yogi, *J. Appl. Phys.* **1990,** *68*, 1169.
[12] H. W. van Kesteren, A. J. den Boef, W. B. Zeper, J. H. H. Spruit, B. A. Jacobs, P. F. Garcia, *J. Appl. Phys.* **1991,** *70*, 2413.
[13] A. Wadas, H. J. Hug, *J. Appl. Phys.* **1992,** *72*, 203.
[14] A. J. den Boef, *Appl. Phys. Lett.* **1990,** *56*, 2045.
[15] A. Wadas, H.-J. Güntherodt, *Phys. Lett. A* **1990,** *146*, 277.
[16] P. Grütter, D. Rugar, H. J. Mamin, G. Castillo, S. E. Lambert, C.-J. Lin, R. M. Valletta, *Appl. Phys. Lett.* **1990,** *57*, 1820.
[17] K. I. Schiffmann, *Nanotechnology* **1993,** *4*, 163.
[18] H. J. Mamin, D. Rugar, J. E. Stern, R. E. Fontana, Jr., P. Kasiraj, *Appl. Phys. Lett.* **1989,** *55*, 318.
[19] D. Rugar, H. J. Mamin, R. Erlandsson, J. E. Stern, B. D. Terris, *Rev. Sci. Instrum.* **1988,** *59*, 2337.
[20] J. R. Barnes, S. J. O'Shea, M. E. Welland, *J. Appl. Phys.* **1994,** *76*, 418.
[21] T. Göddenhenrich, H. Lemke, U. Hartmann, C. Heiden, *Appl. Phys. Lett.* **1990,** *56*, 2578.
[22] J. Moreland, P. Rice, *J. Appl. Phys.* **1991,** *70*, 520.
[23] A. Wadas, H. J. Hug, H.-J. Güntherodt, *Appl. Phys. Lett.* **1992,** *61*, 357.
[24] I. D. Mayergoyz, A. A. Adly, R. D. Gomez, E. R. Burke, *J. Appl. Phys.* **1993,** *73*, 5799.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

4 Ballistic Electron Emission Microscopy

4.1 Introduction

Over the last 30 years or so, the field of surface science has addressed the structural, electronic, and vibrational properties of solid surfaces, adsorbates, and thin films [1]. More recently, the revolutionary invention of the scanning tunneling microscope [2] has impacted surface science in an unprecedented way by providing the scientific community with a tool for probing the physical and chemical properties of surfaces on an atomic scale. The capabilities of atomic-scale imaging with scanning tunneling microscopy (STM) have been coupled with local electron spectroscopy by exploiting the bias dependence of electron tunneling in the vicinity of the scanning tunneling microscope tip [3, 4]. In this regard, STM and scanning tunneling spectroscopy (STS) are extremely valuable techniques for probing both surface geometric and electronic structure as compared to more classical, spatially averaged techniques. Extensions of the principles of STM have led to the development of a host of scanning proximal probe instruments for studying surfaces and interfaces [5, 6], most notably scanning force microscopies (SFMs). These scanning probe microscopies (SPMs) have significantly impacted fields such as biology and electrochemistry which are far removed from the established ultrahigh vacuum (UHV) "classical" surface science environment.

A major issue in surface science continues to be the growth of thin films and the behavior of the interfaces between films and substrates. This field of study is clearly related to electronic device technology. Metal/semiconductor interfaces can exhibit diode characteristics due to the formation of interfacial Schottky barriers [7]. Although the classical model for Schottky barrier formation predicts that the resultant barrier height can be formulated by knowing the semiconductor electron affinity and the metal work function, it has been known for several decades that other factors such as interface states, the doping density of the semiconductor, and interfacial reactions can result in large deviations from the classical model. For instance, it has been demonstrated that the Schottky barrier height can vary for the same two materials grown (epitaxially) with two different orientations at the interface [8]. This provides strong evidence that the roles of local atomic structure and interfacial band structure are very important in determining the Schottky barrier heights at metal/semiconductor interfaces.

In addition, many interfacial systems might not be atomically abrupt over macroscopic dimensions. Therefore, structural inhomogeneities coupled with relatively long screening lengths in the semiconductor play a major role in determining the Schottky barrier height of the resulting device.

A fundamental understanding of hot-electron processes at metal/semiconductor interfaces is important for the optimization of the charge transport properties of many electronic devices such as metal-base transistors and infrared Schottky photodiodes. Experimentally, probing buried interfaces, especially over lateral dimensions where inhomogeneities are likely, is difficult. Standard charged-particle probes do not provide much subsurface information. Nonlinear optical spectroscopy shows some promise in probing buried interfaces, addressing their electronic structure. However, transport measurements across device-size interfaces do not account for local inhomogeneities. STS has contributed to the body of knowledge in thin film technology but only probes the surface electronic structure. STM has provided a great deal of insight into epitaxial and nonepitaxial thin film growth processes by resolving the structure of thin film surfaces. Both of these techniques are limited for interfacial investigations because they are not capable of probing subsurface electronic and geometric structure. The combination of the proximal probe capabilities of STM and carrier transport through an interface describes the unique capability of ballistic electron emission microscopy (BEEM). With this technique, electrons are collected after they traverse a tunnel junction, a metal-base layer, and a semiconductor substrate. Besides information on the local Schottky barrier height, BEEM probes many aspects of hot-electron transport across a heterojunction with the nanometer resolution that is amenable to the study of variations in thin film growth and research on new device technologies that include nanometer structures.

BEEM was developed at the Jet Propulsion Laboratory in 1988 by Kaiser and Bell [9] as an STM-based technique capable of probing the electronic properties of buried metal/semiconductor interfaces with high spatial resolution via the analysis of the transport of hot electrons. BEEM is a three-probe STM-based technique where the tip is placed above a metal/semiconductor heterojunction, which is typically unbiased, and acts as a highly localized variable-energy electron injector. An experimental configuration for BEEM is depicted in Fig. 1. The metal film (thicknesses typically <50 nm), or base, is set at zero potential, as is the substrate, or collector. The tip, or emitter, bias defines the energy distribution of the injected electrons. A comparison of the tip and collector currents (of the order of 1 nA and 10 pA, respectively) as a function of tip bias provides the means to measure the fraction of the electrons that have traversed the metal film and surmounted the Schottky barrier. This describes BEEM, which is also referred to as ballistic electron emission spectroscopy, or BEES. By using a scanning tunneling microscope tip as the injector, the charge injection site is highly localized and can be scanned. Therefore, it is possible to spatially map the potential barrier as a current image. The entire process can, in principle, be modeled theoretically, provided that the electron

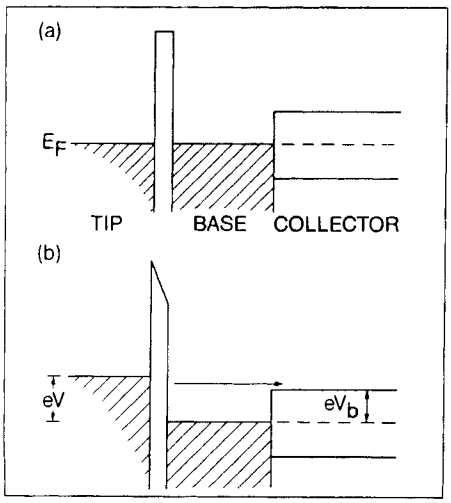


Figure 1. Illustration of the three-terminal BEEM experimental configuration for the local microscopy and spectroscopy of buried metal/semiconductor contacts. The metal overlayer is grounded, the tip bias (V_t) defines the electron injection energy (eV_t), and the unbiased substrate is the collector (a). A BEEM spectrum represents the collector current (I_c) as a function of electron injection energy at constant injection current (I_t) (b). A BEEM image is a collector current map for a specific electron injection energy. STM structural images and BEEM current images of heterojunctions are typically acquired simultaneously to distinguish morphology and interfacial electronic structure. From Ref. [9].

trajectories and scattering and energy loss mechanisms are understood. In the standard BEEM configuration electrons are injected into the metal base and couple to the conduction band states of the semiconductor interface; this requires the use of an n-type substrate that repels the negative charge carriers away from the interface into the semiconductor. Other configurations of BEEM that will be discussed briefly below can probe ballistic hole transport and electron-hole scattering by performing spectroscopy at positive and negative biases and using p-type substrates.

4.2 Experimental Considerations

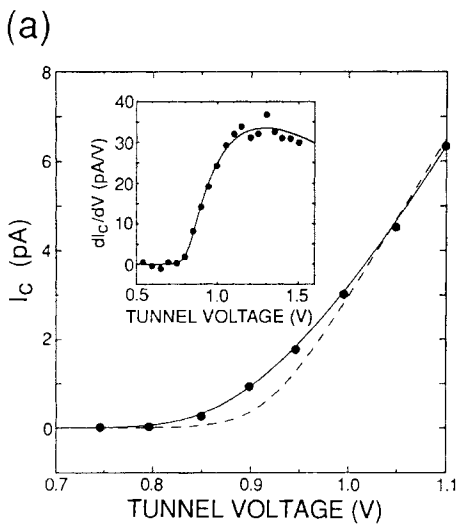
BEEM instrumentation closely resembles that of STM, but several refinements in sample handling, electronics, and vibration isolation are required. As a three-terminal experiment (see Fig. 1), it is necessary to independently contact the front and back of the (unbiased) sample. It is extremely important that the tunneling current (injected electron flux) remains constant, which requires excellent vibration isolation since the tunneling gap is typically controlled by the STM z-axis feedback control (constant current mode). An ohmic contact to the rear of the substrate (the collector), which is often formed by using indium or highly doped silicon, is necessary to allow the collection of the BEEM electrons. Contact to the metal base can be made by a thin wire provided the base contact is continuous. Interfaces with small lateral dimensions (area $\leqslant 1\,\text{mm}^2$) require a base contact very close to the tunneling point, which necessitates micropositioning of this contact [10]. With collector currents typically in the picoampere range, high-gain ($\approx 10^{11}$), low-noise, low-input-impedance preamplifiers are used. One of the most important design criteria is the zero-bias resistance of the diode itself since the voltage noise of the preamplifier will result in noise currents across the junction. With conventional state-of-the-art solid state electronics, typical zero-bias resistances of the base–collector junction should be of the order of $\approx 100\,\text{k}\Omega$ or higher to maintain picoampere resolution. Since the zero-bias resistance is inversely proportional to the diode area and is exponentially dependent on the inverse of the temperature, high

zero-bias resistances can be achieved by either decreasing the diode area or the temperature. For the low barrier heights encountered in many systems ($\leqslant 0.7$ eV), low-temperature operation is necessary since the diode areas required to perform BEEM at room temperature would be impractically small. The latest generation of instruments are proceeding in this direction.

Depending on the information that is desired, there are several ways to configure a BEEM experiment. Thin film interfaces are typically grown using surface science techniques in a UHV chamber. However, just as the STM can be operated in air or a gas as well as in a vacuum, many BEEM experiments can be performed in a controlled medium at atmospheric pressure as long as the metal base resists oxidation. This simplifies low-temperature measurement, which can then be made by immersion in a cryogenic liquid or in a flow of cooled gas. In many cases, the metal base cannot be fully passivated; therefore, a totally in situ experiment is required. Recently, a BEEM system which combined UHV operation with low-temperature capabilities has been designed and demonstrated [11].

4.3 First Demonstrations of Ballistic Electron Emission Microscopy

The first BEEM experiments were performed on Au/Si(100) and Au/GaAs(100) heterojunctions. Those results demonstrated both the subsurface imaging and local spectroscopic capabilities of BEEM [9]. It was previously recognized from macroscopic junction diode measurements that Au/Si exhibits fairly consistent Schottky barrier heights, while Au/GaAs exhibits a greater range of barrier heights, which was interpreted as resulting from interdiffusion and alloy formation in the GaAs system. Figure 2 shows BEEM spectra for these systems [14] where the measurement is localized in the vicinity of the scanning tunneling microscope tip. Here, the degree of localization depends on the electron trajectory through the metal film and interface. These spectra are plots of I_c versus V_t for a constant tunneling current, I_t, and represent, in the simplest interpretation, diode forward-bias characteristics. Based on modeling the threshold spectra, Schottky barrier heights of approximately 0.85 and 1.2 eV for Au/Si(100) and Au/GaAs(100), respectively, were deduced. Figure 3 shows typical STM images and BEEM images for these interfaces [9]. For the BEEM images, the electron energy was set above the threshold and a current image was recorded. Comparing STM topographic images with BEEM current images permits the distinction of morphological and interfacial effects. The Au/Si(100) STM image exhibits a smooth topograph, and the BEEM current image shows a similar uniformity. This result is consistent with the small variations in values of barrier height for Au/Si diodes in general. The surface morphology of gold on GaAs, on the other hand, is considerably rougher, and the BEEM current exhibits considerable inhomogeneity. Furthermore, features in the STM and BEEM images are not correlated, supporting the hypothesis that defect formation and interdiffusion



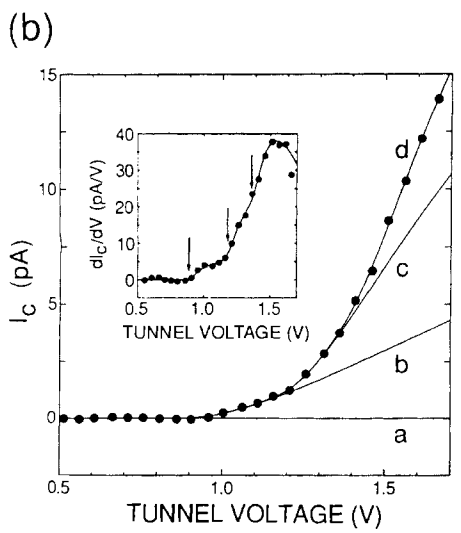


Figure 2. (a) (I_c) versus (V_t) and dI_c/dV_t versus V_t spectra near the Schottky barrier threshold region for an Au/Si(100) heterojunction. A simple one-dimensional theoretical fit (dashed line) was improved (solid line) with the inclusion of parallel momentum conservation at the interface and thermal broadening. (b) I_c versus V_t and dI_c/dV_t versus V_t spectra near the Schottky barrier threshold region for an Au-GaAs(100) heterojunction showing the direct minimum and other band structure-dependent thresholds. From Ref. [14].

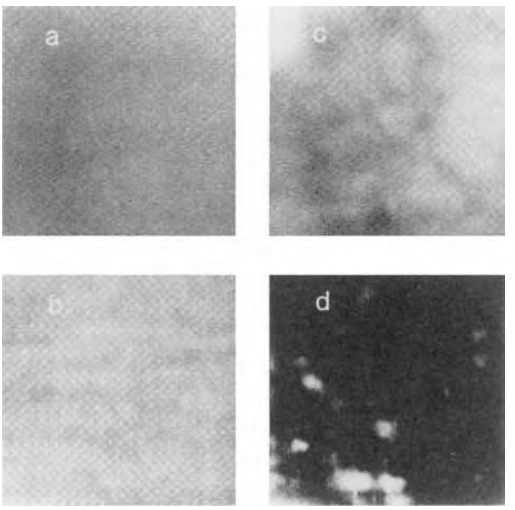


Figure 3. Simultaneously acquired (a) STM and (b) BEEM images ($25.0 \times 25.0\,nm^2$) of an Au/Si(111) heterojunction. The STM (2.2 nm height range) and BEEM (1.5 pA collector current range at $V_t = 1$ V and $I_t = 1$ nA) images indicate that both the metal film structure and the Schottky barrier are very uniform over the collector. Simultaneously acquired (c) STM and (d) BEEM images ($25.0 \times 25.0\,nm^2$) of an Au/GaAs(110) heterojunction. The STM (7.2 nm range) and BEEM (1.5 pA collector current range at $V_t = 1.5$ V and $I_t = 1$ nA) images both show considerable nonuniformity. The fact that topographic and current features are not correlated suggests that multiphase formation creates a nonuniform Schottky barrier. [9].

exist at the interface. Note that topographic features at the 1 nm level could also be seen in the BEEM images. A closer look at the BEEM spectra for Au/GaAs(100) indicates the existence of multiple thresholds. These thresholds occurring at 0.89, 1.18, and 1.36 eV become more dramatic in the first derivative spectrum. They arise from interface band structure assuming that the parallel component of the electron momentum, $k_{||}$, is conserved as the electrons travel across the metal/semiconductor interface.

4.4 Theoretical Considerations

Electron transport in solids can be probed with an unprecedented level of spatial resolution using BEEM. As a direct result of the development of BEEM, there has been a renewed interest in understanding the details of hot-electron transport phenomena in thin-film systems. Analyses of the spectral shape of the BEEM current in the near-threshold region, where much of the transport equation can be simplified, has resulted in very accurate determinations of Schottky barrier heights. However, high-quality fits beyond the threshold require a substantially increased level of complexity. Benchmark experiments have provided insights into various components of the analysis, and at least a qualitative understanding of many novel effects has been achieved. Since the transport problem cannot be solved in its entirety by direct comparison of experimental data to first-principles calculations, many interpretations remain controversial.

The theoretical analysis of a BEEM experiment must account for the physical interactions that govern electron transport from the scanning tunneling microscope tip to the back of the semiconductor substrate. A list of several considerations which must be taken into account to properly characterize the transport of electrons across a Schottky barrier is presented in Table 1. The four major components of BEEM electron transport are: (1) vacuum tunneling of the electrons into the metal base; (2) transport through the thin film; (3) transport across the metal/semiconductor interface; and (4) transport through the semiconductor depletion region. A general expression for the BEEM current at a temperature T can be written as

$$I_b \propto \int D(T,E) f(E,T) E \, dE$$

where $f(E,T)$ is the energy distribution of electrons reaching the interface and $D(E,T)$ is the total transmission probability for electron transport from the point of injection to the back of the semiconductor. In their pioneering BEEM studies of the Au/Si(100) interface, Bell and Kaiser [14] derived an expression for

Table 1. Theoretical considerations for BEEM spectral analysis.

- *Tunneling*
  - Tip
    - Bias (injection energy)
    - Shape/size/angular spread of electrons
    - Tunneling theory
  - Metal thin film surface
    - Surface electronic structure
    - Surface topography/gradients
- *Transport in the film*
  - Metal film (ballistic transport)
    - Growth morphology
    - Thickness
    - Elastic/inelastic scattering
      - Electron–electron
      - Electron–phonon
      - Defects
  - Quantum interference effects
    - Energy dependence
    - Band structure
- *Transport across the interface*
  - Schottky barrier
  - Epitaxy/non-epitaxy/$k_{\|}$ conservation/non-conservation
  - Band-structure match-up
  - Dislocations, defects, etc.
- *Transport in the semiconductor*
  - Image potential effects
  - Dopant density
  - Band structure
  - Elastic/inelastic scattering
    - Electron–electron
    - Electron–phonon
  - Impact ionization (electron–hole pair generation)
  - Depletion region
- *The effects of interlayers*

the bias dependence of the BEEM current by assuming that the energy dependence of both the transmission probability and the energy distribution of the injected electrons can be neglected near the threshold region, which results in a square law dependence for I_b versus V_t given by

$$I_b \propto e^2(V_t - \phi_{SB})^2$$

where ϕ_{SB} is the Schottky barrier height. However, in a BEEM experiment the energy distribution of the injected electrons decays exponentially from a maximum value of eV_t. In addition, the effect of quantum mechanical reflection at the metal/semiconductor interface is strongest near the threshold region. These effects are expected to result in deviations from the square law dependence of the BEEM current within a few hundred millielectronvolts of threshold. The inclusion of the energy dependence of the transmission coefficient due to quantum mechanical reflection results in a 5/2-power law given as

$$I_b \propto e^2(V_t - \phi_{SB})^{5/2}$$

for the expansion of I_b near threshold [12]. On the other hand, it is expected that phonon scattering within the depletion region [13] will partially cancel the effect of quantum mechanical reflection at finite temperature yielding a result which again approaches the square law dependence derived by Bell and Kaiser [14].

To properly interpret BEEM data, a fundamental understanding of the complete electron transport process from injection into the metal overlayer to collection in the semiconductor is necessary. The term *hot electron* is often used to describe electrons in BEEM transport processes. Hot electrons are simply free carriers with kinetic energies greater than approximately kT (40 meV at room temperature). Specifically, these electrons must undergo scattering processes in order to reach thermal equilibrium [15].

The energy and momentum distribution of electrons injected into the metal overlayer has a strong influence on the functional dependence of the spectroscopic lineshape and the interfacial resolution of BEEM. The energy distribution of the injected electrons will decay exponentially, with the actual strength of the decay constant depending both on the height and shape of the vacuum tunneling barrier and the tip apex and surface geometries [16]. Calculations of the energy and momentum distributions of electrons injected into a planar metallic jellium slab from a single-atom scanning tunneling microscope tip apex have indicated that the energy distribution of electrons should be 200 meV or less at energies above 5 eV and that the angular distribution of the electrons should be within about $\pm 10^\circ$ of the surface normal [17]. This forward focusing of the BEEM electrons predicts nanometer-scale spatial resolution at the metal/semiconductor interface for metal overlayer thicknesses of 10.0 nm or less, and such spatial resolution has been observed for many systems. However, for compound metals such as $CoSi_2$ and $NiSi_2$, the influence of the spatial dependence of the local tunneling probability is expected to significantly modify the energy and momentum distribution of the injected electrons. In addition, the atomic potentials of the metals atoms are also expected to affect the actual trajectories of the electrons through the metal base. For ordered metal overlayers, low-energy electron diffraction (LEED) calculations are probably more appropriate for determining the

actual momentum distribution of the electrons as they approach the metal/semiconductor interface.

A determination of the relative length scales of elastic and inelastic scattering processes in the metal overlayer is of great importance since these scattering events have a direct effect on the resolution of BEEM at the interface and limit the range of metal overlayer thicknesses which can be probed. Electrons can scatter elastically off defect sites or quasielastically by emission or absorption of acoustic phonons. For metals with more that one atom per primitive basis, the emission and absorption of optical phonons is also possible. Although the energy quanta of acoustic phonons are typically of the order of a few millielectronvolts, the energies of optical phonons for compound metals, such as metal silicides, can exceed 50 meV. Therefore, the determination of whether electron scattering with optical phonons is a quasielastic process depends on the particular modes available for that system. Hot electrons may also undergo inelastic electron–electron collisions with electrons near the Fermi level, resulting in an average energy loss of one-half of their kinetic energy. When the inelastic mean free path for scattering, λ_i, is of the same order as the elastic mean free path, λ_e, the transport process is essentially *ballistic* since electrons which have undergone multiple scattering events will, on average, not have enough kinetic energy to cross the Schottky barrier. In this limit, I_b should show a clear exponential dependence on metal film thickness with the attenuation length, λ_b, measured by BEEM, given by

$$\frac{1}{\lambda_b} = \frac{1}{\lambda_e} + \frac{1}{\lambda_i}$$

In the limit where $\lambda_e \ll \lambda_i$, the transport is expected to be diffusive in nature since the probability for undergoing multiple elastic scattering events before inelastic scattering is high. This can result in an electron momentum distribution at the metal/semiconductor interface that is completely unrelated to the momentum distribution at the point of injection.

At the metal/semiconductor interface, a fraction of the incident electrons will backscatter into the metal overlayer due to quantum mechanical reflection. For nonepitaxial systems and nonabrupt interfaces, there will be a break in symmetry parallel to the interface, resulting in additional interfacial scattering. For epitaxial interfaces, $k_{||}$ is conserved as the electron travels from the metal to the semiconductor. This condition results in a critical angle for transport into the semiconductor substrate as depicted in Fig. 4. Assuming free electron dispersion relations for the tip and base and a parabolic conduction band minimum for the semiconductor, Bell and Kaiser [14] derived an expression for the

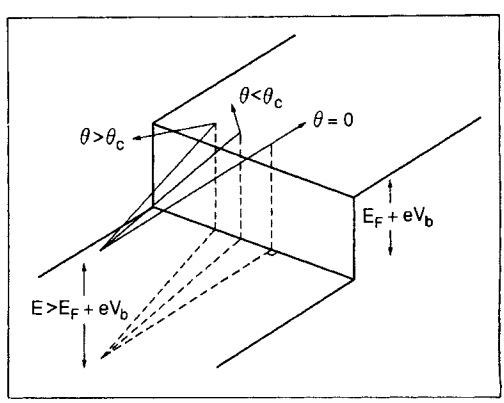


Figure 4. Diagram showing a particle of energy E incident on a potential step of height $E_F + eV_b$. There is a critical angle, θ_c, beyond which the particle is reflected. For angles of incidence, $\theta < \theta_c$, the particle is transmitted with refraction. [37].

critical angle of transmission, θ_c, for the Au/Si(100) system, given as

$$\sin^2 \theta_c = \frac{m_s}{m} \frac{eV_t - e\phi_{sb}}{E_F + eV_t}$$

where E_F is the Fermi energy of the metal, m is the free electron mass, and m_s is the effective mass of the electron in the semiconductor. As a direct consequence of $k_{||}$ conservation, the interfacial spatial resolution of BEEM within a few electronvolts above threshold is expected to exceed the forward-focusing resolution [17]. In addition, a delayed threshold for I_b is expected for electron transport into semiconductor surfaces where the conduction band minimum is not projected onto the surface Brillouin zone center.

Of the key issues which have resulted in conflicting interpretations of BEEM data, the question of whether $k_{||}$ is conserved at the metal/semiconductor interface has remained the most controversial. Although most models for interpreting BEEM electron transport, as well as electron transport from internal photoemission data [18], assume that $k_{||}$ is conserved as the electron crosses the interface, in retrospect this appears to be a somewhat questionable assumption since most metal/semiconductor systems are nonepitaxial in nature. A more relevant issue would be the degree of scattering at nonepitaxial interfaces. In the limit of a jellium metal, the lattice positions of the metal atoms have no influence on the electron trajectory through the metal film, which implies that $k_{||}$ should be conserved. For d-band metals where the band structure deviates rather significantly from that of a simple metal, scattering at the interface might be expected to be almost isotropic. For all real interfaces, there will be crystallographic imperfections and interface reactions which will also contribute to the scattering process. Presently, a systematic approach for accurately evaluating the degree of scattering at metal/semiconductor interfaces for any particular nonepitaxial system does not exist.

Although scattering processes within the semiconductor substrate are not expected to influence the spatial resolution of BEEM at the interface, they will affect the magnitude of I_b. Electrons can scatter by emission or absorption of either acoustic or optical phonons. Due to the image potential [19], the position of the Schottky barrier maximum is not at the metallurgical metal/semiconductor interface but is shifted into the semiconductor by a few nanometers. Electrons with energies just over the threshold for transmission that excite phonons in the region before the Schottky barrier maximum can be expected to have a high probability of reentering the metal. Beyond the Schottky barrier maximum, the internal electric field in the depletion region accelerates the electrons toward the interior of n-type semiconductors. Therefore, the effect of phonon scattering beyond the Schottky barrier maximum on the magnitude of I_b depends on the doping density of the semiconductor, since this defines the length of the depletion region, and thus the acceleration rate. Once the kinetic energy of the electrons exceeds the band gap of the semiconductor, electron–hole pair generation, or impact ionization, becomes possible. Since the internal electric field in the semiconductor will sweep the electrons toward the interior of the semiconductor and the hole toward the metal, an electron multiplication process occurs.

4.5 Ballistic Electron Emission Microscopy Analysis of Schottky Barrier Interfaces

The major emphasis in BEEM analysis is on electron transport across the Schottky barrier interface. The question often comes down to whether or not $k_{||}$ is conserved in transport across the interface. To first order, this can be answered by determining whether the spectral threshold is consistent with the band structures of the metal base and the semiconductor substrate, assuming that $k_{||}$ is conserved. However, scattering in the base is a mechanism that can provide parallel momentum to a fraction of the electrons which reduces the threshold to its minimum value, relaxing band structure considerations. Beyond threshold, other scattering and band structure effects in spectra provide insights into hot-electron transport throughout the junction and electronic states at the interface.

A successful demonstration of the quantitative modeling of measured BEEM spectra in which various components of electron transport could be separated was reported recently [20]. BEEM spectra for gold and magnesium films grown on n-type GaP(110) were fitted over a range of 6 eV as shown in Fig. 5. Spectral decomposition using best-fit parameters distinguishes contributions due to zero scattering (dashed lines), elastic scattering (dashed–double-dotted lines), inelastic scattering plus secondaries (dotted lines), and impact ionization (dashed-dotted lines). One notable difference between these systems is that the overall current magnitudes for the Mg-GaP system are considerably less than that for gold (≈40% of injected current for 1.2 nm-thick gold films), and the spectral shapes are different, especially near threshold. The difference in the total transmission probability between the two films was found to originate from vastly different mean free paths ($\lambda_{in}^0 = 51.0\,\text{nm}\,\text{eV}^2$ for

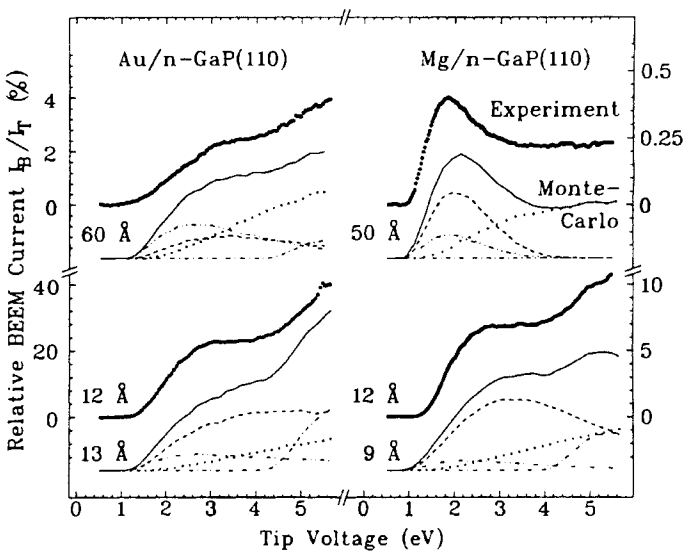


Figure 5. A comparison of measured and simulated BEEM spectra for several film thicknesses of magnesium and gold on n-type GaP(110). Zero scattering (dashed lines), elastic scattering (dashed–double-dotted lines), inelastic scattering plus secondary electrons (dotted lines), and impact ionization (dashed–dotted lines) contributions to the measured spectra are isolated in the theoretical calculations. [20].

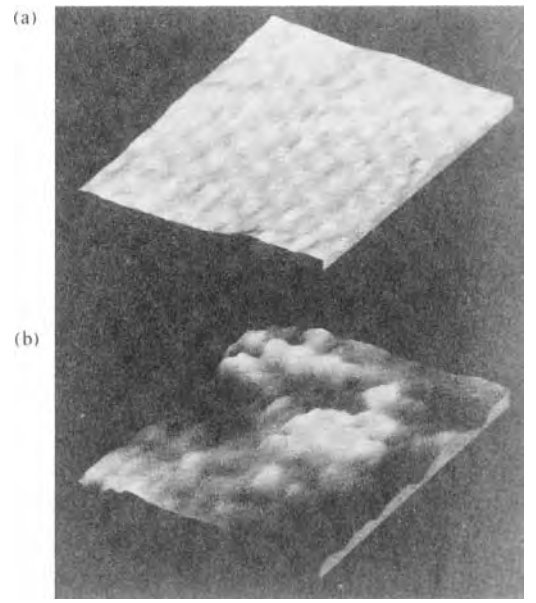
gold and $\lambda_{in}^{0} = 8.5\,\text{nm}\,\text{eV}^2$ for magnesium). At energies below 4 eV the strong inelastic scattering causes an overshoot in the BEEM current for the 5.0 nm-thick magnesium films and obscures impact ionization contributions.

4.5.1 Epitaxial Interfaces

The $CoSi_2$/Si interface is a system under study for applications such as metal base and permeable base transistors. The Schottky barrier properties of both epitaxial $CoSi_2$/Si(111) and $NiSi_2$/Si(111) diodes, extending over macroscopic dimensions, have been studied intensively in recent years since it was determined that the barrier height could be controlled as a function of growth conditions [8]. These cubic fluorite silicides can be grown with two relative lattice orientations on Si(111), forming high-quality atomically abrupt interfaces which account for the barrier height differences. Since both systems are epitaxial, it is expected that a $k_{||}$ model for electron transport across the interface is appropriate since there is no break in symmetry parallel to the interface.

For electrons with $k_{||}$ precisely equal to zero, the energy band alignment results in a gap approximately 1.4 eV above E_F for the $CoSi_2$/Si(111) interface. Calculations for the $CoSi_2$/Si(111) interface predict that transmission can occur about 0.2 eV above the conduction band minimum [21]. Measurements on $CoSi_2$ films grown on n-type Si(111) and imaged in liquid nitrogen indicated typical threshold values at 0.85 eV [22] at many locations and on different samples. However, several locations gave threshold values of about 0.65 eV, which is the approximate value for the Schottky barrier height, with some difference in spectral shapes. Examples of the measured spectra are shown in Fig. 6. In addition, the BEEM current images exhibit considerable inhomogeneity at the interface which correlates with the range of threshold values. Such observations of BEEM thresholds above the Schottky barrier were the first such results to be reported. These results, including their unique spectral shapes, indicate that the electron transport measurements contain details of the band structure of the epitaxial interface when modeled with the inclusion of $k_{||}$ conservation. Although the previous experimental results seem to provide evidence for $k_{||}$ conservation for electron transport across an epitaxial interface, recent BEEM results for $CoSi_2$/Si(111) measured in situ at 77 K give evidence to the contrary [11]. BEEM current thresholds of 0.66 eV were measured for all regions of their samples with no additional current onset at 0.85 eV. The discrepancies between these two studies is most likely a result of the different sample preparation and measurement conditions. The lack of a delayed threshold for the in situ study indicates that an additional source of scattering at the metal/semiconductor interface was present.

BEEM studies on the $NiSi_2$/Si(111) system, where $k_{||}$ conservation is also expected, are more difficult to interpret since the band structure does not provide an energy gap but rather an overlap between $NiSi_2$ and Si states at all energies above the Schottky barrier [23]. $NiSi_2$ films of various thicknesses were grown on an Si(111)-7 × 7 interface and analyzed in situ [24]. Atomically resolved STM images of the silicide surface were compared with the



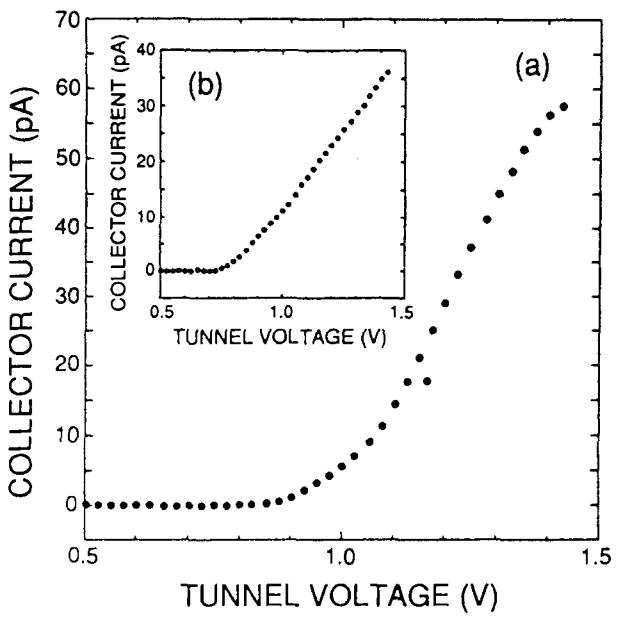


Figure 6. *Upper:* simultaneously acquired STM (a) 2.1 nm height range and BEEM (b) 40 pA collector current range at $V_t = 0.5$ V and $I_t = 1$ nA images ($25 \times 25\,nm^2$) of a 10 nm-thick $CoSi_2$ film on n-type Si(111). *Lower:* a representative BEEM spectrum for the $CoSi_2$ film most commonly observed at various sample positions. The inset shows representative spectra that are less commonly observed. [22].

7×7 unit cell of the bare Si(111) surface to determine the orientation of the silicon films. The thinner $NiSi_2$ films (<2.5 nm) were B-type (rotated 180° with respect to the substrate) whereas the thicker films were A-type and B-type mixed interfaces. BEEM spectra were acquired on $NiSi_2$ terraces as well as in pinholes where the 7×7 surface is viewed as a zero-thickness metal overlayer. On terraces, transmission of forward-focused electrons cannot occur without a minimum $k_{||}$. Thus, the

signature of $k_{||}$ conservation was taken to be the observed soft threshold, which is also predicted by theory [23]. This observation was contrasted with a sharp threshold in the pinhole regions which is attributed to disorder. In addition, a 25% increase in the BEEM current was observed in the region immediately surrounding the $NiSi_2$ terraces. This provides additional evidence for $k_{||}$ conservation because the angular distribution of the injected electrons at the terrace edges is expected to be broadened, thus resulting in increased coupling to silicon conduction band states located away from the surface Brillouin zone center.

4.5.2 Nonepitaxial Interfaces

A study of electron transport in thin films where $k_{||}$ conservation was expected to break down was performed on palladium films of various thicknesses (0.8–9 nm) grown on Si(111) and Si(100) substrates [10]. The surface topography as measured by STM provided an assessment of local surface gradients which were small enough such that off-normal injection would not lead to direct transmission into conduction band minima at $k_{||} \neq 0$. Figure 7a shows a series of BEEM spectra as a function of bias. The threshold is at approximately 0.73 eV, independent of the thickness, and the spectra are linear to approximately 2 eV, at which point there is a fall-off due to the density of states effects. A second threshold, most evident for the thinnest films, appears at approximately 2.9 eV and arises from electron–hole pair creation. For thicker films, the second threshold was less intense due to scattering in the

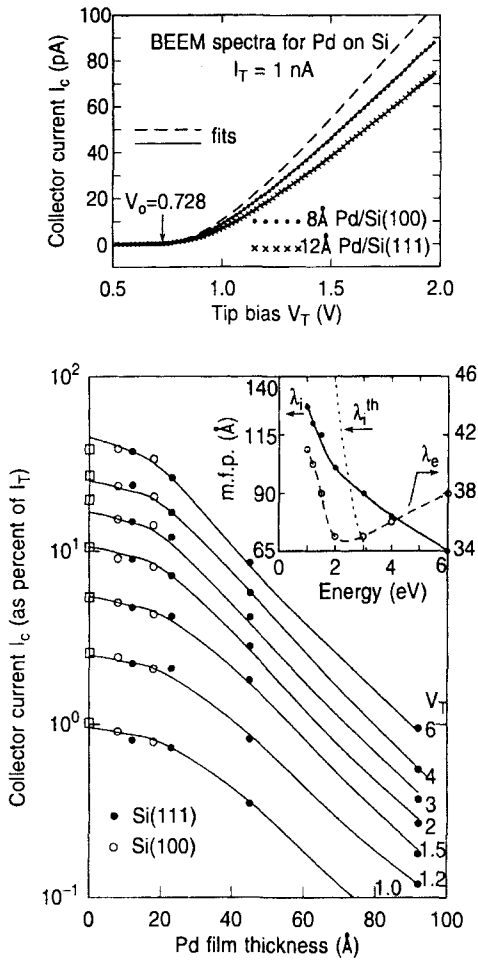


Figure 7. (a) Comparison of BEEM spectra for 0.8 nm Pd/Si(100) and 1.2 nm Pd/Si(111) illustrating the differences due to scattering at the interface. The dashed curve was generated for zero metal thickness and a transmission probability of 0.7. (b) Attenuation of I_c with palladium thickness as a function of tip bias, V_t. The inset shows the energy dependence of the elastic and inelastic electron mean free paths. [24].

metal base layer. A plot of collector current versus palladium film thickness is shown in Fig. 7b; both Si(111) and Si(100) substrate data lie along the same respective curves as expected for nonconservation of $k_{||}$. The attenuation of the BEEM current exhibits two slopes due to the fact that the

scattering in thin metals is less significant than in thicker metal films. A model based on an isotropic momentum distribution at the interface was developed to analyze the series of plots of attenuation behavior. This result (inset to Fig. 7b) gave the energy-dependent elastic and inelastic mean free paths (λ_e and λ_i, respectively) which were fitted from the data using the phenomenological equation

$$p = \exp\frac{-w}{\lambda_b} + \left(1 - \exp\frac{-w}{\lambda_e}\right)\exp\frac{-4w^2}{\lambda_e\lambda_i}$$

where p is the probability of an injected electron reaching the interface and w is the film thickness. The first term is the probability of surviving ballistic passage and the second term is the product of the probability of elastic scattering and the probability of surviving $\bar{n} \approx 2w/\lambda_e$ scattering events without inelastic scattering. According to this formulation, both the elastic and inelastic mean free paths can be treated independently.

4.5.3 Au/Si Interfaces

The Au/Si system was the first used to demonstrate the BEEM technique, and is probably one of the most comprehensively studied systems. However, there remains a curious set of inconsistencies in the data regarding these interfaces, some of which can be directed toward variations in the interfaces themselves and some in sample preparation. Several investigators have observed BEEM current for these interfaces while others have measured no current. Some investigators have concluded the validity of $k_{||}$ while others have discounted this assumption.

It has already been noted that the pioneering experiments of Kaiser and Bell demonstrated very good agreement between their data and the model assuming $k_{||}$ conservation at the Au/Si(100) interface [14]. However, similar lineshapes for Au/Si(100) and Au/Si(111) have been observed [25], which is inconsistent with $k_{||}$ conservation because of band structure differences between the two substrate faces. In essence, transport into Si(111) requires a finite $k_{||}$ to couple to the states at the conduction band minimum while transport into Si(100) does not. Their Monte Carlo calculations suggested that strong elastic (and quasielastic) scattering in the gold creates an isotropic momentum distribution which reduces the spatial resolution and relaxes the requirement of $k_{||}$ conservation at the interface. This conclusion was further supported by BEEM measurements which observed no correlation between surface gradients and BEEM current [26]. In another study of Au/Si, an SiO_2 interlayer grid was grown at the interface to provide information on the limits of subsurface spatial resolution [27]. It was found that BEEM currents varied over distances of about 1.0 nm, leading to a conclusion that ballistic transport must dominate through the gold layer. It was also noted that the observed spatial resolution was significantly higher than would be expected from simple planar tunneling theory. High interfacial resolution of BEEM for Au/Si(100) interfaces was also confirmed in a study in which Schottky barrier height fluctuations were directly imaged with nanometer-scale resolution for highly doped substrates [28]. However, another study found that no BEEM current was transmitted through an Au/Si(111) contact that was grown

under UHV conditions [29]. This result was attributed to a reaction at the interface forming a disordered gold–silicon alloy, which acts as a source of strong scattering. Growing gold on an H-terminated surface resulting in uniform BEEM currents. In addition, it was noted that higher biases caused an irreversible modification in BEEM transmission for these interfaces, which was attributed to an electron-induced enhancement of gold–silicon interdiffusion [29]. More recently, however, Cuberes et al. [30] found that BEEM currents could be measured up to 8 eV tip biases without resulting in interfacial modification for UHV-grown samples. Trying to reconcile all these results is difficult because there are several experimental differences, especially in sample preparation and growth and in methods of detection.

It appears that electron scattering mechanisms and the scattering strength in the metal layer are important components for electron transport analysis. Recently, a temperature-dependent BEEM study of the Au/Si(100) interface was performed for various layer thicknesses ($\approx$6.5–34.0 nm) over a narrow range of energies [31]. The attenuation length in the gold overlayer was determined to be energy-independent with only a small temperature dependence leading to the conclusion that defect scattering is the dominant scattering mechanism in the metal film. The temperature-dependent electron transmission through the semiconductor depletion region was also probed and found to be dominated by acoustic phonon scattering. These measurements indicated that, at room temperature, almost half of the electrons which pass into the semiconductor are backscattered into the metal base.

4.5.4 Metal-Film Dependence

Classical Schottky barrier theory predicts that the barrier height is dependent on the macroscopic properties of the metal and semiconductor. This theory breaks down when the actual atomic structures of interfaces of highly epitaxial systems are taken into account [8]. It has also been shown that structural inhomogeneities account for major variations in Schottky barrier heights [32]. Even without a detailed analysis of individual interfacial structures, it is still expected that different metals on the same substrate should produce different Schottky barrier heights. A systematic BEEM investigation of magnesium, silver, copper, nickel, and bismuth on GaP(110) was performed in a UHV BEEM system where the films were grown and analyzed in situ [33]. A different threshold was observed for each metal film as shown in Fig. 8, and differences in the average electron transmission coefficients were noted for the different metals. Variations in the BEEM current at different positions on the sample were attributed to the surface morphology of the respective film (see below). Despite such current variations, the threshold positions were maintained, demonstrating the uniformity of the interfacial potential barrier.

4.5.5 Surface Gradients

In the systematic study of various metals grown on GaP(110) described above, it was found that the surface morphology is the major determinant for current variations in a BEEM image—not interfacial structure, which might also be present and provide contrast in a BEEM image [33]. Figure 9

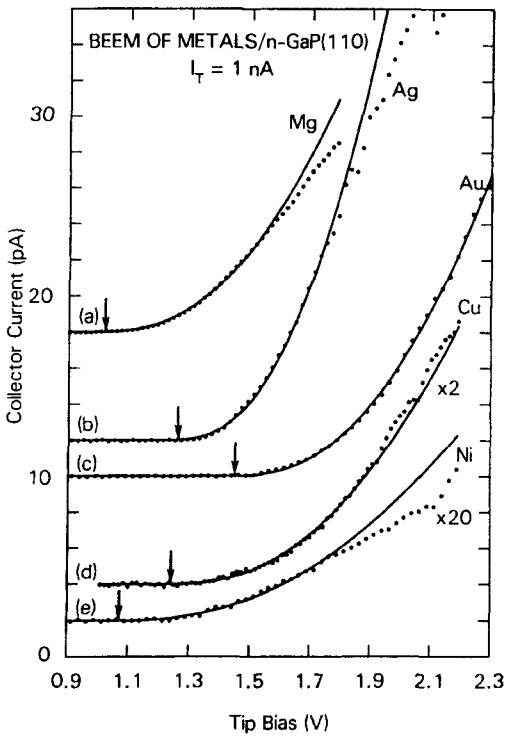


Figure 8. BEEM spectra of various metal/GaP(110) interfaces: 8.0 nm Mg, 20.0 nm Ag, 6.0 nm Au, 15.0 nm Cu, and 5.0 nm Ni. The energies of Schottky barrier thresholds are found to depend on specific metal overlayers, while spectra taken at various positions for a given sample exhibited fixed thresholds. Ref. [33].

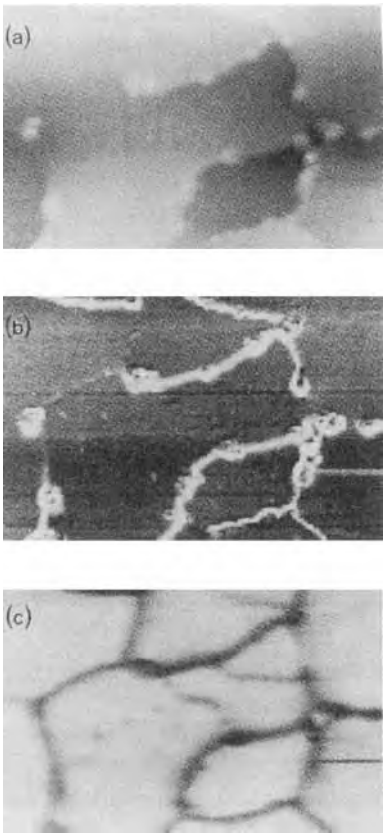


Figure 9. (a) Topographic image (1.2 nm height range) of a 60.0 × 40.0 nm^2 area of a 20.0 nm thick magnesium film on GaP(110); (b) gradient image of (a) (0–30% gradient range); (c) BEEM image (8–28 pA collector current range) of (a) collected simultaneously. Ref. [33].

compares the surface morphology of a 20.0 nm magnesium film with its respective BEEM image. In the BEEM current image, the lowest currents were recorded beneath positions on the surface with high topographic gradients. A closer look at the surface morphology, where inclinations of below 10° existed, led to the conclusion that the transmitted current is reduced because the electrons are injected into the metal film off-normal (with respect to the interface). Neglecting scattering in the film, the forward momentum component is no longer normal to the interface, which results in reduced coupling to the substrate band structure and can result in total reflection away from the barrier, causing a marked reduction in collector current. The effect of surface gradients on the measured BEEM current is commonly referred to as the "searchlight effect".

4.5.6 Interfacial Nanostructures

Perfectly epitaxial, atomically abrupt interfaces are idealizations of real

heterojunctions. While considerable efforts are made to prevent interfacial inhomogeneities such as disorder, defects, dislocations, and other nanoscale structures at interfaces, the specific material properties often conspire against ideal structures. As noted above, interfacial imperfections represent scattering centers that destroy the periodicity and $k_{||}$ conservation at the interface. Besides, such features can result in variations in the local Schottky barrier height. It is reasonable that interfacial defects might create sharp contrast in BEEM images so that the underlying interfacial defect structure can be mapped.

Scattering at interfacial dislocations has been observed for partially strain-relaxed $CoSi_2$ epitaxial films (2–7 nm thicknesses) grown on Si(111) [11, 34]. Film growth was performed in a UHV BEEM system, and imaging was performed in situ at 77 K. The STM topograph of Fig. 10 reveals steps from misorientation of the wafer as well as a hexagonal array of dislocations which result from relaxation of the strain at the interface and produce strain fields that distort the surface to such an extent that they are resolved with the scanning tunneling microscope. The corresponding BEEM images exhibit a current enhancement of approximately 20% at the dislocations

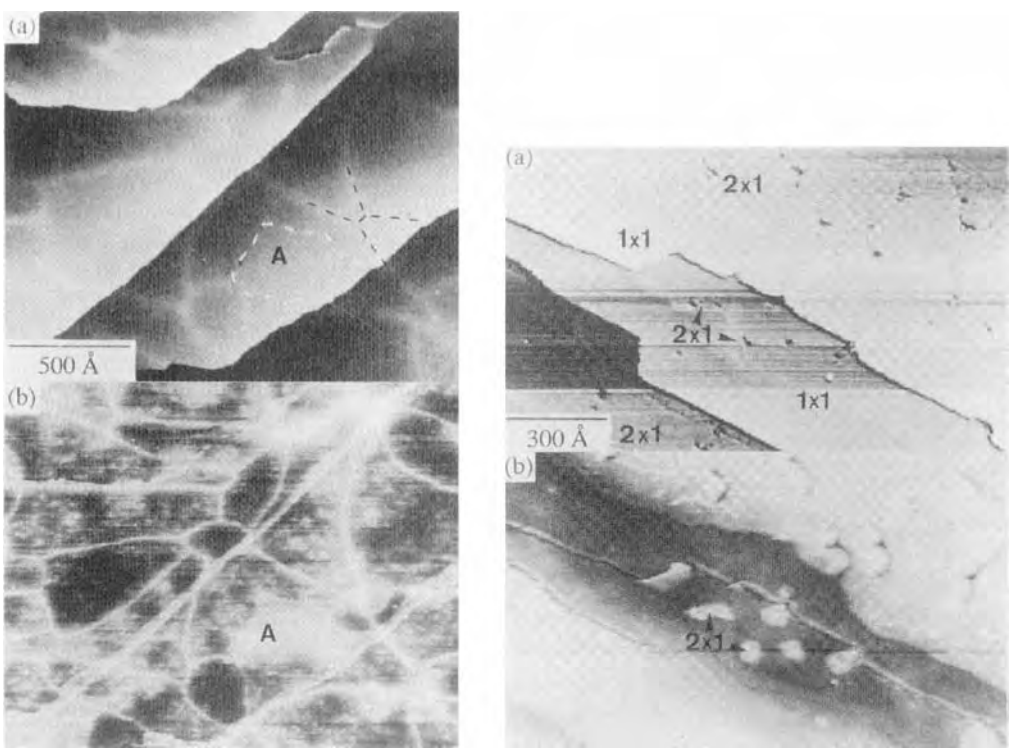


Figure 10. *Left:* (a) STM and (b) BEEM images in the vicinity of dislocations on a 2.5 nm-thick $CoSi_2$/Si(111) heterojunction indicate the variability in BEEM current due to dislocations. Region A exhibits a 2×1 reconstruction, and the arrow is directed at a surface point defect. *Right:* (a) STM and (b) BEEM images on a 1.9 nm-thick $CoSi_2$/Si(111) film. The BEEM current in the 2×1 region is approximately 40% higher than for the 1×1 region. [34].

that decorate the BEEM image. The spatial dimension of the current enhancements at these interfacial imperfections is extremely sharp (≈0.8 nm wide). This implies that the majority of scattering occurs at the interface, which broadens the momentum distribution and enhances the current since access to states at the conduction band minimum requires momentum transfer parallel to the interface. The enhancement was most pronounced approximately 1 eV above the Schottky barrier threshold, whereupon it begins to decrease due to band structure effects. In addition, local defects unrelated to the dislocations were observed in the current images since they also serve as scattering centers which provide parallel momentum transfer. The unstressed interface is not reconstructed. However, region A exhibits a 2×1 reconstructed region that is created by local stress; this region shows an increased current due to a change of surface electronic structure alignment of the conduction band minimum. Small differences in contrast from step to step occur because of quantum interference effects arising from atomic variations in the film thicknesses.

4.5.7 Local Electron Tunneling Effects

Much of the effort in the analysis of a BEEM experiment is devoted to understanding the details of electron transport *after* the electrons have been injected into the metal film. It is clear, however, that the surface electronic structure controls tunneling between the tip and surface atoms. It is recognized that atomic resolution can be achieved in topographic and current STM images and that the local density of states has an effect on the bias dependence of STM images. Recent experiments conducted in a UHV BEEM apparatus at 77 K on $CoSi_2$ grown on both Si(100) ($\sqrt{2} \times \sqrt{2}$R45° and $3\sqrt{2} \times \sqrt{2}$R45° structures) and Si(111) (1×1 and 2×1 structures) surfaces have demonstrated that atomically resolved features can be observed in BEEM images [35]. On n-type substrates, this effect is manifest as a variation of the BEEM current on the atomic scale beyond the Schottky barrier threshold. Figure 11 compares STM and BEEM images for the ($\sqrt{2} \times \sqrt{2}$R45°) structure on Si(100); the BEEM current

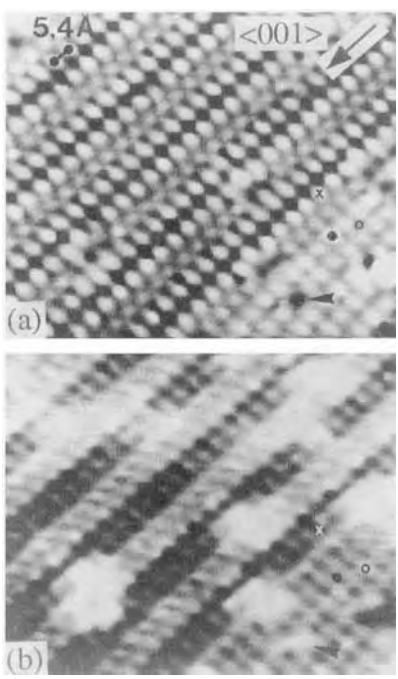


Figure 11. (a) STM topograph and (b) simultaneously acquired BEEM image for the silicon-rich $CoSi_2$/n-Si(100) surface (3.8 nm film thickness). The central area exhibits a $3\sqrt{2} \times \sqrt{2}$R45° reconstruction and the lower right shows a $\sqrt{2} \times \sqrt{2}$R45° reconstruction. [35].

is reduced by approximately 20% at points of topographic maxima over a range of V_t between -1.0 and -2.4 V. This out-of-phase image contrast is similar to that observed for $(3\sqrt{2} \times \sqrt{2}R45°)$. $(\sqrt{2} \times \sqrt{2}R45°)$ and $(3\sqrt{2} \times \sqrt{2}R45°)$ regions can be distinguished even after atomic resolution is lost up to approximately 6 eV because of differences in the average local tunneling barrier heights. The fact that the atomic corrugation directly corresponds to the simultaneously acquired STM image is indicative that some detail of the tunneling interaction at the surface controls the process. Assuming $k_{||}$ conservation at the epitaxial interface, atomic resolution in BEEM could originate from variations in the energy distribution (details of the local tunneling barrier) or the momentum distribution (spread of the angular cone) of the carriers. The difference between injection into Si(100) and Si(111) surfaces is that the projected conduction band minimum is in the center of the Brillouin zone for Si(100). Thus, for the Si(100) substrate, a broadening in the momentum distribution would cause a reduction in the BEEM current especially near the Schottky barrier threshold. The fact that the contrast persists for higher energies, where the momentum distribution is less critical, makes variations in the energy distribution more likely. Furthermore, out-of-phase contrast between STM and BEEM are also seen for Si(111), where a broader momentum distribution should increase the current. At least for energies near threshold, changes in the energy distribution of injected carriers induced by local variations of the tunneling barrier are thus regarded as the origin of atomically resolved BEEM. However, model calculations of transport including the energy and momentum effects still significantly underestimate the observed contrast, so that only a qualitative explanation of atomically resolved BEEM presently exists.

4.5.8 Impact Ionization

An energy loss mechanism for electrons in the semiconductor substrate with a kinetic energy greater than the band gap is the generation of electron–hole pairs. In a BEEM spectrum, this is manifest in the observation of excess current, i.e. electron multiplication effects in the semiconductor, since the internal field in the semiconductor will sweep the holes toward the metal base and the electrons to the back of the depletion region. This phenomenon was first observed for disordered 5.0 nm chromium films that were grown on n-type GaP(110) [36]. More recently, impact ionization effects in BEEM spectra on very thin $NiSi_2$ films grown on Si(111) were reported [24]. BEEM spectra acquired over a high-quality epitaxial region (B) as well as at a pinhole (A), where a modified 7×7 structure exists are shown in Fig. 12. The pinhole spectrum allows a direct analysis of scattering events in the semiconductor since the 7×7 reconstruction results in an ultrathin metallic overlayer. This is regarded as the zero-metal-thickness limit for electron transport. Direct comparison with experiment indicated that primary electrons pass to the semiconductor with a 72% efficiency. Schottky barrier thresholds at 0.79 eV were observed for both the 7×7 pinhole regions and the B-type $NiSi_2$. For injection

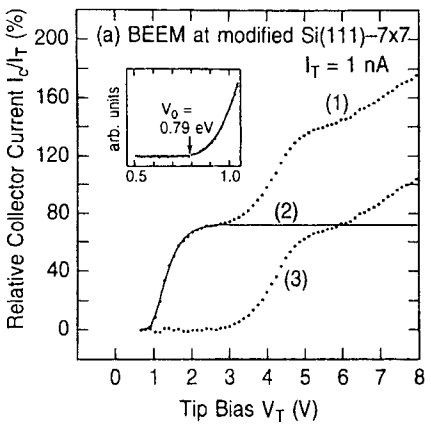


Figure 12. Curve (1) is a representative BEEM spectrum acquired in a pinhole of a 2.2 nm B-type $NiSi_2$ film. Curve (2) is the primary electron component, and curve (3) is the secondary electron component which occurs due to impact ionization in the semiconductor. The inset shows the threshold region of the spectrum. [24].

energies greater than 3 eV, a second current threshold was observed, which is attributed to the onset of impact ionization in the silicon. For kinetic energies of approximately 5 eV, the quantum yield for impact ionization, γ, is of the order of unity, i.e. one primary electron creates a secondary electron.

4.6 Probing Beneath the Schottky Barrier

Thin film systems are not restricted to situations in which an individual layer is deposited on a substrate. Ultrathin multiple-layer structures are particularly relevant for new device applications. As an aid in the evaluation of novel devices, BEEM can be applied as a localized probe of interfacial and thin film electronic properties of semiconducting or insulating interlayers by analyzing electrons that have traversed several layers. To perform the BEEM measurements, electrons are injected into a metal layer that has been deposited on the desired multiple-layer structure. Epitaxy and disorder, band structure and alignment, and electron transport across each interface and through each material can substantially complicate a quantitative BEEM analysis. However, some examples of recent work illustrate the extension of BEEM capabilities to such structures of increased complexity.

Typically, analysis of electron transport in the semiconductor using BEEM is complicated by scattering events in the metal overlayer and at the metal/semiconductor interface which result in modifications of the initial momentum and energy distributions of the injected electrons. However, a heterostructure configuration—a gold thin film grown on an epitaxial p-type layer on an n-type Si(100) substrate—has been used recently to address electron scattering effects in the semiconductor [48]. From these results temperature-dependent attenuation lengths within the semiconductor were determined. With a p–n configuration beneath the Schottky barrier, the actual barrier maximum can be placed tens of nanometers below the Schottky barrier. The barrier maximum is expected to be dependent on the thickness. When the p layer is thick enough so that it is not fully depleted, the barrier maximum depends only on the doping density. In addition, the barrier maximum is expected to be temperature-dependent due both to temperature effects in electron scattering and to the band structure. Experimental measurements were made for p-type layers with thicknesses of 0–100 nm, and

theoretical modeling, incorporating $k_{||}$ conservation at the interface, produced excellent fits to the data. The barrier height saturated at an approximate thickness of 40 nm thickness because the layer was not fully depleted, and the temperature shift in the barrier height was in agreement with previous data. Interestingly, the data analysis produced an unexpected result regarding the temperature dependence of the attenuation length, $\lambda(T)$. In particular, the value of approximately 45.0 nm obtained for $\lambda(300\,\mathrm{K})$ is significantly longer than that the valued obtained for $\lambda(77\,\mathrm{K})$ of 15.0 nm. This was tentatively attributed to the need for increased scattering at higher temperatures to provide electron momentum to couple the injected electrons to the states at the silicon conduction band minimum that exist $k_{||} \neq 0$.

The measurement of conduction band discontinuities and transport across flat-band semiconductor heterostructures beneath the Schottky barrier has also been demonstrated. Conduction band offsets for $Al_xGa_{1-x}As/GaAs$ heterostructures as a function x (the aluminum mole fraction) were investigated [38]. The thin film configuration shown in the inset of Fig. 13a was adopted for these measurements. A p-type δ-doped layer (beryllium) of appropriate density was inserted during growth to compensate for the band bending due to the Schottky barrier (at 300 K). The measured spectra, obtained both at 300 and 77 K for $x = 0$, 0.21, and 0.42, are shown in Fig. 13b. The $x = 0$ threshold (0.92 eV) is consistent with previously determined Schottky barrier values for Au/GaAs. The observed trend, an upshift in threshold energy with increasing x, can be attributed to the energy barrier of the AlGaAs interlayer where this barrier

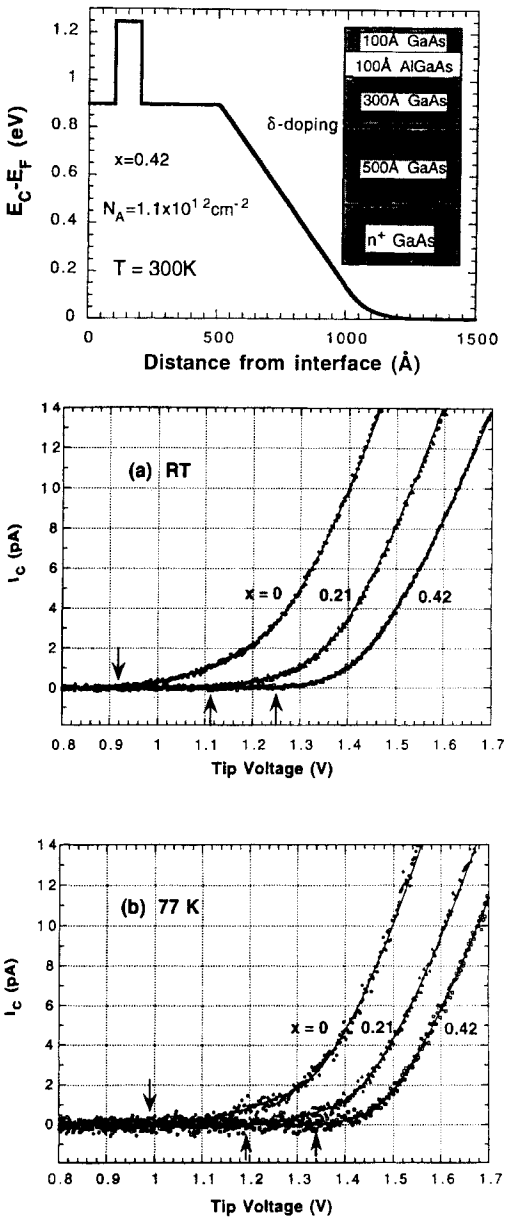


Figure 13. *Upper:* calculated conduction band energy profile for the Γ valley of $Al_{0.42}Ga_{0.58}As$ on a GaAs structure assuming a Schottky barrier height of 0.9 eV at room temperature. A p-type δ-doping sheet is indicated. *Lower:* comparison of BEEM spectra for $Al_xGa_{1-x}As$ /GaAs single-barrier structures (at RT (a) and 77 K (b)). Changes in the threshold voltages correspond to band offsets of the GaAs and AlGaAs Γ conduction bands. [36].

height is the difference between thresholds. Two-threshold fits in accordance with the respective band structures (Γ and L for $x = 0$ and Γ and X for $x = 0.21, 0.42$) were found to best model the spectra for the determination of a precise threshold value. Although these offsets could be measured by other techniques, the fact that the BEEM measurement is localized laterally might aid in the optimization of ultrathin structures with confined (e.g. nanometer scale) lateral dimensions.

Metal/insulator/semiconductor (MIS) structures are used in many current electronic devices such as MOSFETs. Interest has recently focused on the application of CaF_2 as insulating thin films because of a small lattice mismatch with silicon offering the potential for epitaxial growth. The geometric and electronic structure of CaF_2 on Si(111) has been studied by a host of techniques including STM [39]. A BEEM study of the Au/CaF_2/Si(111) MIS structure focused on determining the effect of a wide band gap insulating layer on electron transport [30, 40]. The most significant observation was that BEEM spectra conform to the general width of the CaF_2 density of states, that is the conduction band states "filter" the ballistic electrons that traverse the film. In addition, the CaF_2 interlayer was found to inhibit the propagation of surface silicides which exist for gold growth on bare silicon at room temperature. Figure 14 compares the spectra of Au/Si(111) and Au/CaF_2/Si(111). For Au/Si(111), the BEEM threshold appears at 0.75 eV above E_F, and a feature at 3 eV was attributed to the onset of impact ionization. When a CaF_2 interlayer of four monolayers thickness (≈1.4 nm) was incorporated, the current threshold shifted by approximately

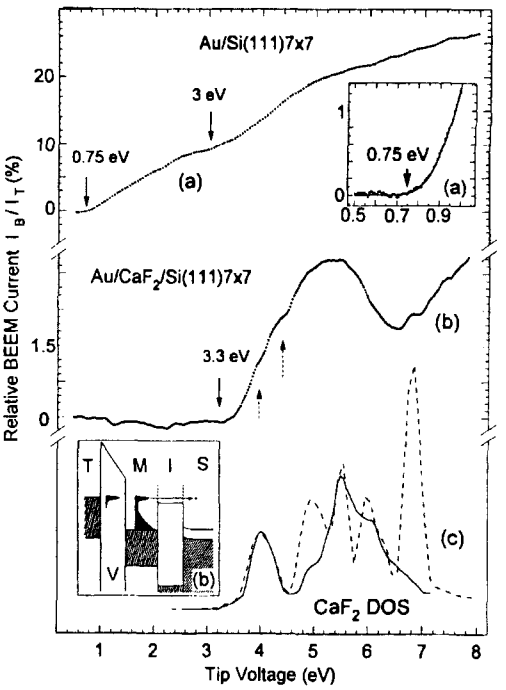


Figure 14. Representative BEEM spectra for (a) a 30 monolayer gold film on Si(111)–7 × 7 and (b) a 30 monolayer gold film plus four monolayer CaF_2 film on Si(111)–7 × 7. The inset shows the energetics of the experiment for an MIS structure. (c) The calculated bulk CaF_2 density of states. [40]..

2.5 eV to about 3.3 eV above E_F. In addition, the current diminished approximately 3 eV above this threshold, forming an energy window which is approximately the width of the CaF_2 density of conduction states, which is also shown in Fig. 14. Small features marked by arrows relate to possible impact ionization, but other spectra features can also be related to peaks in the density of states. It should be noted that the density of states in Fig. 14 must be projected out in the direction of electron transport for any direct comparison of features. In addition, the rising background which is observed for higher energies is expected and is due to inelastic processes in electron transport.

Although SiO_2 is used as a gate oxide in virtually all production MOSFETs, little experimental data exist on the hot-electron properties across these thin films on an atomic scale. This is primarily due to the amorphous nature of the SiO_2 in these devices which limits the analysis of their structural and transport properties to experimental techniques which do not rely on crystalline order. Recently, an in situ BEEM study of the $Pt/SiO_2/Si(100)$ MOS structure was performed to determine directly the phonon scattering rates of electrons through the conduction band of SiO_2 with nanometer spatial resolution [41]. The samples were prepared by evaporating platinum onto device quality oxide layers approximately 6.2 nm thick which were thermally grown on n-type Si(100) wafers. In addition to the conventional BEEM set-up, a variable bias was applied to the semiconductor substrate with respect to the grounded metal base. This set-up allowed field-dependent transport measurements through the SiO_2 conduction band and also provided the means to tune the energy distribution of electrons that pass into the silicon substrate. A rapid rise in I_c was observed at a threshold of approximately 4 eV for $V_b = 0$ V and +2 V followed by an overshoot in I_c at approximately 6 eV and a more gradual rise above 8 eV, which can be attributed to the onset of electron–hole pair generation in the silicon substrate. This attenuation has been attributed to the strong electron–phonon coupling in SiO_2 since the scattered electrons are increasingly accelerated toward the metal base for decreasing values of V_b. The measured attenuation rates are in good agreement with previous Monte Carlo calculations of electron–phonon interactions in SiO_2 [42], and reinforce the interpretation that hot-electron scattering with the acoustic phonon modes dominate over the optical modes in SiO_2 in this energy range.

A key consideration in the development of real electronic devices is the stability of the interface in its operating environment. Au/GaAs(100) interfaces degrade at room temperature due to gold diffusion into the surface. It was demonstrated that spatial Schottky barrier inhomogeneities exist at oxide-free Au/GaAs interfaces directly after growth [9] (see Fig. 3). BEEM experiments have also shown that an AlAs interlayer inserted as a diffusion barrier between the gold and GaAs greatly improves both the homogeneity of the Schottky barrier and the stability of the device [43]. Surface science experiments typically strive to obtain atomically clean surfaces prior to deposition. However, in the fabrication of GaAs-based devices, an oxide is typically left behind, which results in the enhancement of device stability. BEEM has been used to investigate the effect of a native oxide interlayer over time on the electron transport properties of Au/GaAs(100) interfaces [44]. An approximately 2 nm-thick native oxide ($Ga_2O_3 + As_2O_3$) on GaAs was left prior to gold deposition. Figure 15 shows that BEEM spectra are still obtainable after more than a month at room temperature in air. However, it was found that many regions did not support a BEEM current after this long a period, suggesting that the oxide slows but does not arrest chemical reactions at the interface. Figure 15 also shows that with increasing time the zero-bias resistance of the interface decreases, resulting in an increase of the current below the threshold bias. This was attributed to diffusion of gold atoms into

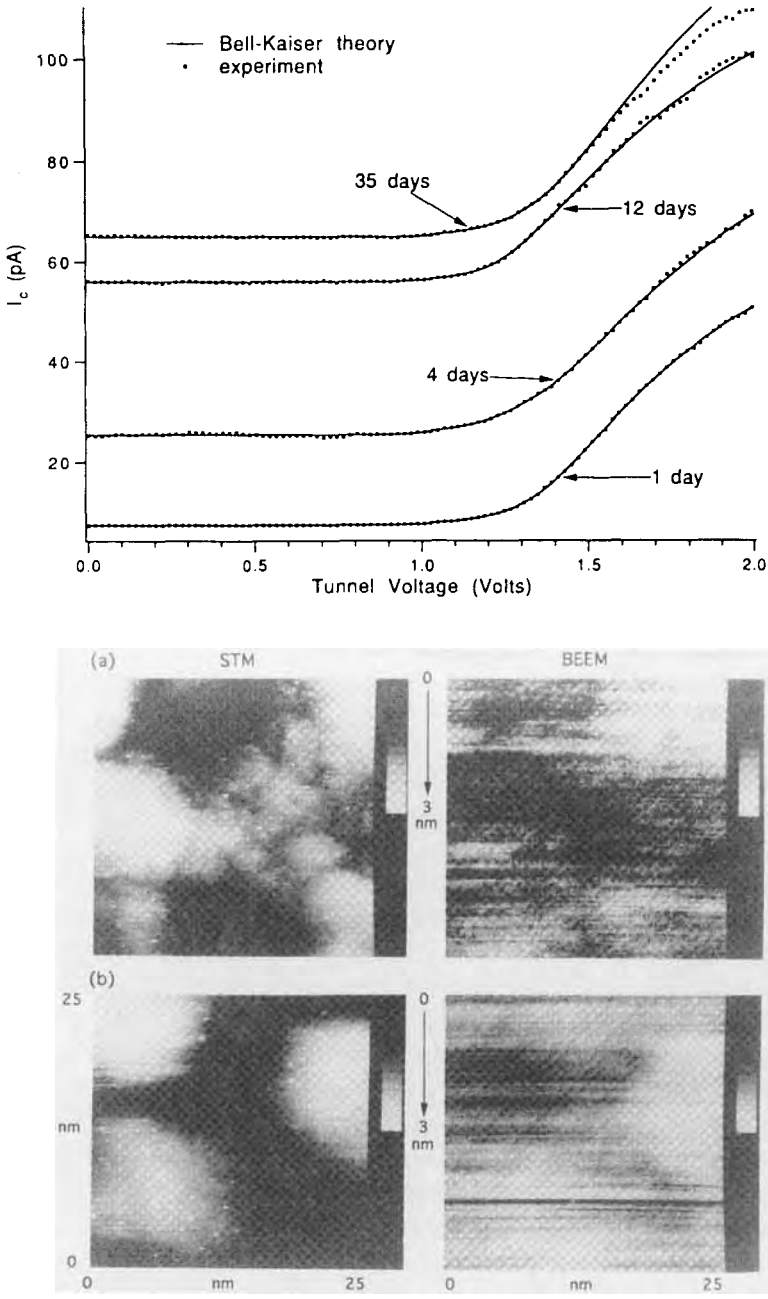


Figure 15. *Upper:* average BEEM spectra acquired with $I_t = 2$ nA obtained at different locations on an Au/n-GaAs(100) sample at various times after deposition (the 35 day spectrum is displaced upward by 5 pA). *Lower:* STM and BEEM images acquired (a) 1 day after deposition and (b) 12 days after deposition. [44].

the GaAs, displacing arsenic and acting as dopants or recombination centers. The native oxide exhibits a zero or nearly zero band gap and does not cause significant electron scattering at the interface. This is in contrast to metals grown on thin oxide layers on silicon. These layers possess a significantly large band gap that impedes

the ballistic electron current across the interface [45].

4.7 Ballistic Hole Transport and Ballistic Carrier Spectroscopy

The standard BEEM configuration probes semiconductor conduction band states at the interface. This results from the injection of electrons into the metal base and their subsequent passage into the collector, which is dependent on the energetic position of the conduction band states of the semiconductor. An n-type semiconductor substrate is used to prevent majority carriers from entering the base.

Shortly after the introduction of the BEEM technique, the Jet Propulsion Laboratory group demonstrated a methodology with which to probe the semiconductor valence band states using ballistically transported holes [46, 47]. For this, the metal base was deposited on a p-type substrate, and hole injection was accomplished by reversing the electron tunneling direction with respect to BEEM on an n-type substrate, that is, from base to tip. An additional observation was made that electron and hole scattering spectroscopy in the base region can be achieved by performing $I-V$ spectroscopy over both polarities for either metal/n-type or metal/p-type junctions.

Figure 16 shows a schematic that depicts ballistic hole transport and electron–hole scattering and measured spectra for Au/Si(100) systems acquired at low temperature. A low temperature was required because of the magnitude of the Schottky barrier heights. We first consider the ballistic electron and ballistic hole spectra. For BEEM on n-type substrates, electrons are transported across the Schottky barrier (tip negative), and the conduction band levels of the metal/semiconductor interface are probed. For BEEM on p-type substrates, holes are transported across the Schottky barrier (tip positive), and the valence levels of the metal/semiconductor interface are probed. There is an asymmetry in the carrier distributions that are probed. With n-type substrates, the highest electronic levels of the tip are probed; with p-type substrates, the lower levels of the hole distribution take part in conduction. For these respective cases, the number of electrons created per unit energy increases with increasing bias, or the number of holes created per unit energy decreases with increasing bias. Therefore the current threshold for electron transport is at the position of the conduction band minimum, and the threshold for hole transport is at the valence band maximum. The region between these thresholds gives a direct measure of the semiconductor energy gap.

Reversing the bias for either of these cases probes electron and hole scattering in the base. The respective spectra correspond to those carriers that have undergone scattering events and can be modeled as an internal Auger process. For example, in the BEEM configuration with n-type semiconductor substrates, electrons cannot be extracted from the substrate because the Schottky barrier potential opposes the motion of majority carriers into the base. However, a bias configuration where electrons are extracted from the

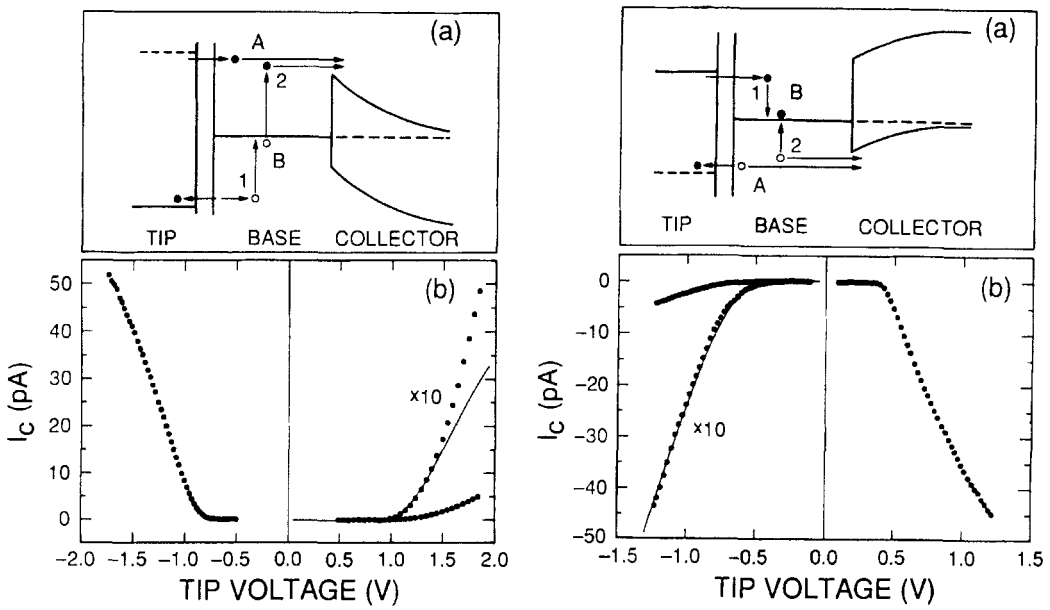


Figure 16. *Left:* (a) The energetics of electron transport for a metal base on an n-type collector (A) using a negative tip voltage and for scattering (B) where the tip is biased positive. Process B creates a ballistic hole distribution that scatters with equilibrium electrons to produce hot electrons. (b) Spectra for Au/Si exhibit an onset at 0.82 eV, the Schottky barrier height for this n-type sample. The scattering spectrum is seen to be less intense by a factor of approximately 10. *Right:* (a) The energetics of electron transport for a metal base on a p-type collector (A) using a positive tip voltage and for scattering (B) where the tip is biased negative. For situation B, electrons injected by tunneling produce a distribution of hot holes after scattering with equilibrium electrons in the base. (b) Spectra for Au/Si exhibit an onset at 0.35 eV, the Schottky barrier height for this p-type sample. Again, the scattering spectrum is seen to be less intense by a factor of approximately 10. Combining the results for these two samples, a gap equal to the substrate band gap is obtained. [46].

base (tip positive) creates energetic holes in the metal base. Thus, ballistic holes scatter from equilibrium electrons and can excite hot electrons above the Schottky barrier. These electrons are collected following transmission across the barrier. Note that the sense of the current is the same for biasing in both polarities. For p-type semiconductor substrates, ballistic electron scattering and hot-hole creation are the analogous processes that occur. In scattering spectroscopy, the currents are attenuated by a factor of approximately 10 from the respective direct transport measurement, indicating an efficient electron-scattering process. The theoretical fits shown in Fig. 16 are valid only around the threshold.

Ballistic carrier transport and scattering can be modeled using the BEEM formalism where momentum and energy conservation are primary considerations. Energy loss mechanisms showing fully inelastic losses for the promotion of electrons and holes and isotropic scattering angular distributions are assumed in the model that was used to fit the data in Fig. 16. It is interesting to note that comparison of ballistic hole scattering for Au/Si(100) and Au/GaAs(100) reveals that the threshold

shape for the GaAs substrate must take into account the markedly larger spin–orbit splitting of hole bands near the zone boundary [47].

4.8 Summary

Methods for the measurement and analysis of local Schottky barrier heights and electron transport in thin film interfaces has developed rapidly in the years since BEEM was introduced. Although the use of the BEEM technique is still restricted to relatively few groups, instrumentation is becoming more sophisticated, and interest in the technique is expanding toward applications on real devices. While first-principles BEEM calculations are on the horizon, comprehensive theoretical calculations continue to give way to more phenomenological analyses which provide new insights into transport phenomena in very thin layers and at metal/semiconductor junctions. The usefulness of BEEM is exemplified by its ability to probe the degree of homogeneity and local Schottky barrier heights by a fairly straightforward measurement—one that could not be done prior to the invention of BEEM—and connected with macroscopic measurements of metal/semiconductor junctions.

Acknowledgments

NJD acknowledges support for this work from the National Science Foundation under grant DMR 93-13047 and via the Materials Research Laboratory Program under grant DMR 91-20668.

4.9 References

[1] C. B. Duke (Ed.), *Surface Science: The First Thirty Years*, **1994,** *300*.
[2] G. Binnig, H. Rohrer, *Surf. Sci.* **1993,** *126*, 236.
[3] R. J. Hamers, R. M. Tromp, J. E. Demuth, *Phys. Rev. Lett.* **1986,** *56*, 1972.
[4] J. A. Stroscio, R. M. Feenstra, A. P. Fein, *Phys. Rev. Lett.* **1986,** *57*, 2579.
[5] N. J. DiNardo, *Nanoscale Characterization of Surfaces and Interfaces*, VCH, Weinheim **1994**.
[6] R. Wiesendanger, *Scanning Probe Microscopy and Spectroscopy: Methods and Applications*, Cambridge University Press, Cambridge **1994**.
[7] W. Schottky, *Naturwissenschaften* **1938,** *26*, 843.
[8] R. T. Tung, *Phys. Rev. Lett.* **1984,** *52*, 461.
[9] W. J. Kaiser, L. D. Bell, *Phys. Rev. Lett.* **1988,** *60*, 1406.
[10] A. R. Ludeke, A. Bauer, *Phys. Rev. Lett.* **1993,** *71*, 1760; b. R. Ludeke, A. Bauer, *J. Vac. Sci. Technol. A* **1994,** *12*, 1910.
[11] H. Sirringhaus, E. Y. Lee, H. von Känel, *J. Vac. Sci. Technol. B* **1994,** *12*, 2629.
[12] M. Prietsch, R. Ludeke, *Surf. Sci.* **1991,** *251/252*, 413.
[13] E. Y. Lee, L. J. Schowalter, *J. Appl. Phys.* **1991,** *70*, 2156.
[14] L. D. Bell, W. J. Kaiser, *Phys. Rev. Lett.* **1988,** *61*, 2368.
[15] C. R. Crowell, S. M. Sze, *Solid-State Electron.* **1965,** *8*, 979.
[16] J. Tersoff, D. R. Hamann, *Phys. Rev. Lett.* **1983,** *50*, 1998.
[17] N. D. Lang, A. Yacoby, Y. Imry, *Phys. Rev. Lett.* **1989,** *63*, 1499.
[18] C. R. Crowell, S. M. Sze in *Physics of Thin Films* (Eds.: G. Hass, R. F. Thun), Academic Press, New York **1967**, pp. 325–391.
[19] S. M. Sze, *Physics of Semiconductor Devices*, Wiley, New York **1984**.
[20] A. Bauer, M. T. Cuberes, M. Prietsch, G. Kaindl, *Phys. Rev. Lett.* **1993,** *71*, 149.
[21] M. D. Stiles, D. R. Hamann, *J. Vac. Sci. Technol. B* **1991,** *9*, 2394.
[22] W. J. Kaiser, M. H. Hecht, R. W. Fathauer, L. D. Bell, E. Y. Lee, L. C. Davis, *Phys. Rev. B* **1991,** *44*, 6546.
[23] M. D. Stiles, D. R. Hamann, *Phys. Rev. Lett.* **1991,** *66*, 3179.
[24] a. A. Bauer, R. Ludeke, *Phys. Rev. Lett.* **1994,** *72*, 298; b. A. Bauer, R. Ludeke, *J. Vac. Sci. Technol. B* **1994,** *12*, 2667.
[25] L. J. Schowalter, E. Y. Lee, *Phys. Rev. B* **1991,** *43*, 9308.

[26] E. Y. Lee, B. R. Turner, L. J. Schowalter, J. R. Jimenez, *J. Vac. Sci. Technol. B* **1993,** *11*, 1579.
[27] A. M. Milliken, S. J. Manion, W. J. Kaiser, L. D. Bell, M. H. Hecht, *Phys. Rev. B* **1992,** *46*, 12826.
[28] H. Palm, M. Arbes, M. Schulz, *Phys. Rev. Lett.* **1993,** *71*, 2224.
[29] A. Fernandez, H. D. Hallen, T. Huang, R. A. Buhrman, J. Silcox, *Appl. Phys. Lett.* **1990,** *57*, 2826.
[30] M. T. Cuberes, A. Bauer, H. J. Wen, D. Vandré, M. Prietsch, G. Kaindl, *J. Vac. Sci. Technol. B* **1994,** *12*, 2422.
[31] C. A., Ventrice Jr., V. P. LaBella, G. Ramaswamy, H. P. Yu, L. J. Schowalter, *Phys. Rev. B* **1996,** 53.
[32] J. P. Sullivan, *J. Vac. Sci. Technol. B* **1993,** *11*, 1564.
[33] M. Prietsch, R. Ludeke, *Phys. Rev. Lett.* **1991,** *66*, 2511.
[34] H. Sirringhaus, E. Y. Lee, H. von Känel, *Phys. Rev. Lett.* **1994,** *73*, 577.
[35] H. Sirringhaus, E. Y. Lee, H. von Känel, *Phys. Rev. Lett.* **1995,** *74*, 3999.
[36] R. Ludeke, *Phys. Rev. Lett.* **1993,** *70*, 214.
[37] L. D. Bell, W. J. Kaiser, M. H. Hecht, L. C. Davis, in *Scanning Tunneling Microscopy* (Eds.: J. A. Stroscio, W. J. Kaiser), Academic Press, San Diego **1993**, pp. 307–348.
[38] a. J. J. O'Shea, T. Sajoto, S. Bhargava, D. Leonard, M. A. Chin, V. Narayanamurti, *J. Vac. Sci. Technol. B* **1994,** *12*, 2625; b. T. Sajoto, J. J. O'Shea, S. Bhargara, D. Leonard, M. A. Chin, V. Narayanamurti, *Phys. Rev. lett.* **1995,** *74*, 3427.
[39] P. Avouris, R. Wolkow, *Appl. Phys. Lett.* **1989,** *55*, 1074.
[40] M. T. Cuberes, A. Bauer, H. J. Wen, M. Prietsch, G. Kaindl, *Appl. Phys. Lett.* **1994,** *64*, 2300.
[41] R. Ludeke, A. Bauer, E. Cartier, *Appl. Phys. Lett.* **1995,** *66*, 730.
[42] M. V. Fischetti, D. J. DiMaria, S. D. Bronson, T. N. Theis, J. R. Kirtley, *Phys. Rev. B* **1985,** *31*, 8124.
[43] M. H. Hecht, L. D. Bell, W. J. Kaiser, F. J. Grunthaner, *Appl. Phys. Lett.* **1989,** *55*, 780.
[44] A. A. Talin, D. A. A. Ohlberg, R. S. Williams, P. Sullivan, I. Koutselas, B. Williams, K. L. Kavanagh, *Appl. Phys. Lett.* **1993,** *62*, 2965.
[45] H. D. Hallen, A. Fernandez, T. Huang, R. A. Buhrman, J. Silcox, *J. Vac. Sci. Technol. B* **1991,** *9*, 585.
[46] L. D. Bell, M. H. Hecht, W. J. Kaiser, L. C. Davis, *Phys. Rev. Lett.* **1990,** *64*, 2679.
[47] M. H. Hecht, L. D. Bell, W. J. Kaiser, L. C. Davis, *Phys. Rev. B* **1990,** *42*, 7663.
[48] L. D. Bell, S. J. Manion, M. H. Hecht, W. J. Kaiser, R. W. Fathauer, A. M. Milliken, *Phys. Rev. B* **1993,** *48*, 5712.

Part VIII

Image Recording Handling and Processing

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

1 Image Recording in Microscopy

1.1 Introduction

A microscope as a highly quantitative measuring device for objects of small dimensions requires efficient means of recording the output information not only for visual observation but increasingly even for evaluation by computers. The two-dimensional distribution representing the image must be converted into signals to be stored and processed taking into account different points of view. Off-line methods such as photographic recording requiring a long processing time are increasingly being replaced by on-line converters using advanced optoelectronic technology which is much faster, frequently more accurate, and ready for immediate digitizing. The efficiency and accuracy of such devices are governed by quantum processes, and their design needs a careful analysis in order to optimize the performance with respect to the task.

In this chapter the position-sensitive detectors used in fixed-beam microscopes are treated, excluding the scanning methods. The general physical fundamentals of image characterization before and after recording are reviewed, as well as the different quantum conversion effects utilized in various optoelectronic components, and the way to achieve optimum performance which photon and electron image recorders have in common. Parameters such as the detective quantum efficiency, resolution, dynamic range, linearity, dependence on the signal rate, and processing speed, which may have different levels of importance in the various methods of microscopy, are discussed. Subsequently, the general relations will be applied to the different types of radiation and recording systems.

1.2 Fundamentals

1.2.1 The Primary Image

In the image plane of any microscopic system a two-dimensional wave function ('primary image') is formed, the amplitude of which is registered by the recording device. Due to the quantum nature of the radiation the image suffers fluctuations ('noise') governed by Poisson statistics, according to which the probability distribution of finding N quanta in an arbitrary local and temporal measuring interval is

$$P(N, \bar{N}) = \frac{\bar{N}^N}{N!} \exp(-\bar{N}) \qquad (1)$$

resulting in the well-known relation

$\operatorname{var} N = \bar{N}$. As a consequence, in order to detect a small image detail of area d^2 and contrast $C = \Delta n/n$ against a background of the local quantum density n, the signal-to-noise ratio $\mathrm{SNR} = \Delta n d^2/\sqrt{nd^2}$ must be sufficiently large. If a detectability condition $\mathrm{SNR} \geqslant 5$ is assumed, then the quantum density should satisfy the condition [1]

$$n \geqslant \left(\frac{5}{Cd}\right)^2 \tag{2}$$

Thus, the size d and contrast C of detectable small details with weak contrast are limited by the storage limit of the recorder, even if an ideal detector such as a two-dimensional quantum counter without any further restriction of resolution is assumed. Approaching the storage limit, the SNR decreases. Consequently, any discussion of image recording devices must include the storage medium [2].

1.2.2 The General Recorder

A real recorder may be modeled by a two-dimensional device converting the primary image to a type of information to be accumulated in an integrated or subsequent frame store until the required SNR is obtained. The goal of retrieving the full information input may be obstructed by a number of restrictions and disturbances:

(a) the saturation limit of the storage, as mentioned above;
(b) the limited spatial field together with the local sampling intervals, frequently called 'pixels' (picture elements);
(c) local blurring of the image by the point spread function (PSF) of the converter;
(d) additional sources of noise, reducing the SNR to values below that in the primary image.

A useful parameter for discussing such influences is the detective quantum efficiency [3]

$$\mathrm{DQE} := \gamma^2 \frac{\operatorname{var} S_{\mathrm{in}}}{\operatorname{var} S_{\mathrm{out}}} \tag{3}$$

where S_{out} and S_{in} denote the signals at the output and input, respectively, accumulated over a partial field and some measuring time, and $\gamma := \mathrm{d}\bar{S}_{\mathrm{out}}/\mathrm{d}\bar{S}_{\mathrm{in}}$ the slope of the characteristic curve $\bar{S}_{\mathrm{out}}(\bar{S}_{\mathrm{in}})$. In general, the DQE is not only a function of the exposure level (e.g., if the storage limit is approached), it depends also on the size of the chosen partial field if this is small enough to become comparable to the extension of the PSF of the converter. For the time being, we ignore these complications by assuming a sufficiently large field and proportionality between $\bar{S}_{\mathrm{out}}$ and $\bar{S}_{\mathrm{in}}$, and obtain

$$\begin{aligned}\mathrm{DQE} &= \left(\frac{\bar{S}_{\mathrm{out}}}{\bar{S}_{\mathrm{in}}}\right)^2 \frac{\operatorname{var} S_{\mathrm{in}}}{\operatorname{var} S_{\mathrm{out}}} \\ &= \left(\frac{\mathrm{SNR}_{\mathrm{out}}}{\mathrm{SNR}_{\mathrm{in}}}\right)^2 = \frac{\mathrm{SNR}_{\mathrm{out}}^2}{\bar{N}}\end{aligned} \tag{4}$$

which shows the sense of the DQE definition. While an ideal image recorder has $\mathrm{DQE} = 1$, real recorders are characterized by $\mathrm{DQE} < 1$, requiring an exposure $\bar{N}$ increasing with DQE^{-1} in order to display the same SNR. Thus, condition (2) must be replaced by

$$n\,\mathrm{DQE} \geqslant \left(\frac{5}{Cd}\right)^2 \tag{5}$$

This condition is of importance if the number of radiation quanta must be

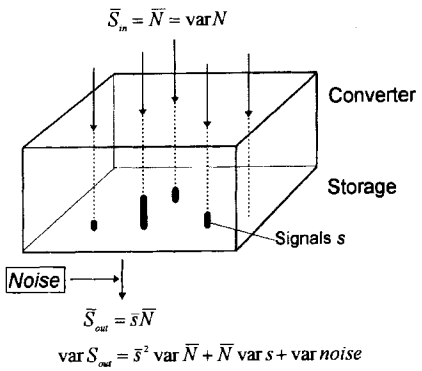


Figure 1. Signal response of an image recorder.

limited, as in imaging radiation-sensitive specimens or dynamic processes using a restricted quantum flux. In these applications particularly, the DQE should approach the ideal value of 1 as closely as possible. Unfortunately, this goal is not always achievable due to the interaction mechanisms which require compromises with the resolution, as discussed below.

An understanding of the degrading statistical processes in an image converter and the design of optimized systems may be facilitated by discussing the signal response to single primary quanta (Fig. 1). Without referring to special carriers of information, this response may be presented by an amount s of secondary quanta (e.g., electrons, photons, or grains) spread over an interaction range within the PSF of the system and fluctuating according to a probability distribution $p(s)$. Under this assumption and using the variance theorem [4], we have $\bar{S}_{\text{out}} = \bar{N}\bar{s}$ and $\text{var}\,S_{\text{out}} = \bar{s}^2\,\text{var}\,N + \bar{N}\,\text{var}\,s$; thus, Eq. (4) may be rewritten as

$$\text{DQE} = \left(1 + \frac{\text{var}\,s}{\bar{s}^2}\right)^{-1} = \frac{\bar{s}^2}{\overline{s^2}} \tag{6}$$

Thus the DQE is independent of the exposure level $\bar{N}$. Obviously, in order to approach the ideal detector the designer should attempt to realize a distribution $p(s)$ as narrow as possible. Unfortunately, this is not always easy, particularly if the quantum energy of the primary radiation is of the order of the formation energy of the secondary quanta.

While the noise phenomena discussed above were related to the signal to be detected, an additive noise contribution by system components may occur which is independent of the signal. Referring it to the output plane and integrating it over the same area and time as the primary quanta, its variance must be added to $\text{var}\,S_{\text{out}}$. Assuming that the noise signal makes no contribution to $\bar{S}_{\text{out}}$ one obtains

$$\text{DQE} = \left(1 + \frac{\text{var}\,s}{\bar{s}^2} + \frac{\text{var noise}}{\bar{N}\bar{s}^2}\right)^{-1} \tag{7}$$

Obviously, this noise contribution makes the DQE dependent on $\bar{N}$ at low exposure. The designer should reduce such system immanent noise as much as possible. The remaining influence may be reduced by choosing $\bar{s}$ large enough to make $\bar{s}^2 \gg (\text{var noise})$. In this case, however, caution is recommended, since then a given limit of an analog storage medium may reduce the DQE at high exposure according to Eq. (3). Thus, the DQE of any recording device fades at both too low and too high exposure levels $\bar{N}$ (Fig. 2); an intermediate working region, the 'dynamic range', with a maximum DQE determined by $\text{var}\,s/\bar{s}^2$, can be found. Its width is determined by the storage limit and the noise, and its position on the exposure scale by the signal height $\bar{s}$.

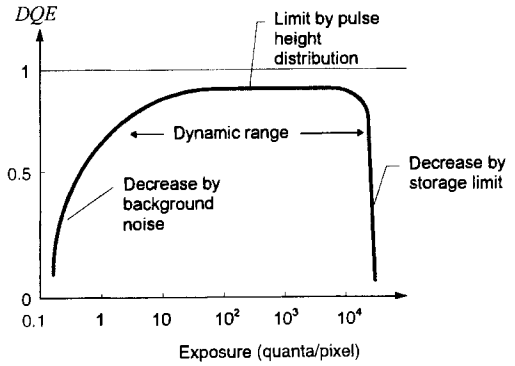


Figure 2. DQE versus exposure of a general recorder shown schematically.

A sufficiently high single-quantum response $\bar{s}$ with low variance may be achieved using one or more quantum conversion stages between electrons and photons as components of imaging optoelectronic devices, as discussed below (Fig. 3). Every conversion process at the stage m $(m = 1, 2, \ldots, n)$ is characterized by a probability distribution $p_m(g_m)$ from which a quantum yield $\bar{g}_m$ and $\operatorname{var} g_m$ can be deduced. By repeated application of the variance theorem [4] it can be shown [5] that the resulting signal s has an average value

$$\bar{s} = \bar{g}_1 \bar{g}_2 \bar{g}_3 \cdots \bar{g}_n \tag{8}$$

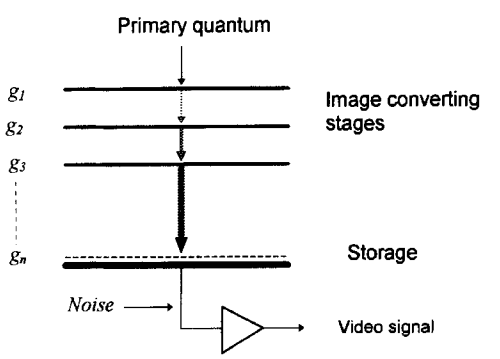


Figure 3. Principle of the multistage image converter.

and a statistical fluctuation

$$\frac{\operatorname{var} s}{\bar{s}^2} = \frac{\operatorname{var} g_1}{\bar{g}_1^2} + \frac{1}{\bar{g}_1}\frac{\operatorname{var} g_2}{\bar{g}_2^2} + \cdots + \frac{1}{\bar{g}_1 \bar{g}_2 \cdots \bar{g}_{n-1}}\frac{\operatorname{var} g_n}{\bar{g}_n^2} \tag{9}$$

From this expression an extremely important design rule can be concluded: the statistical contributions of the subsequent stages with $m > 1$ may be neglected if the products

$$\prod_{i=1}^{m} \bar{g}_i \gg 1 \tag{10}$$

for all stages $m = 1, 2, \ldots, n-1$ (even if some of the quantum yields $\bar{g}_m$ are smaller than unity). A logarithmic 'quantum level diagram' helps to check whether condition (10) is satisfied [6]. If it is, then the first conversion stage dominates the fluctuation $\operatorname{var} s$; consequently, its statistics requires particular attention. It is strongly determined by the interaction process of the primaries with a solid, and frequently co-determined by a compromise between high gain $\bar{g}_1$ and resolution. If the first stage yields $g_1 = 1$ with a probability $p_1(1) = \eta < 1$, and a significant fraction $1 - \eta$ of impinging quanta yields $g_1 = 0$, that is, $p_1(0) = 1 - \eta$, then we have a binary distribution with $\bar{g}_1 = \eta$, $\operatorname{var} g_1 = \eta(1 - \eta)$, and

$$\mathrm{DQE} = \eta\left(1 + \frac{\operatorname{var} s}{\bar{s}^2} + \frac{\operatorname{var}\,\text{noise}}{\eta \bar{N} \bar{s}^2}\right)^{-1} \tag{11}$$

where $\bar{s}$ and $\operatorname{var} s$ refer to the combined yields of all subsequent stages $m > 1$. Thus, η appears as an upper limit of the DQE which cannot be exceeded but only reduced by the variances of the yield s and the noise.

A converting chain is much more meaningfully characterized by the DQE than by a sensitivity factor defined as the ratio

$\bar{S}_{out}/\bar{S}_{in}$ of the means of the output and input, as frequently used, since a degradation of the SNR due to any misdesigned stage violating condition (10) cannot be repaired by postamplification.

Occasionally, a distinct improvement of the DQE can be achieved by a normalization procedure as used in the conventional single-channel radiation-measuring technique. Using a discriminator, the weak and strong noise signals are suppressed and the passing signal pulses are replaced by counting pulses to be stored spatially related into a digital memory. For this method Eq. (11) simplifies to DQE $= \eta$, where η denotes the probability of producing a countable pulse by the primary. This counting technique, introduced by astrophysicists [7], and applied to the three types of primary quanta under discussion, requires some electronic effort in order to determine the 'center of gravity' of the signal spots within the sampling structure and to avoid spatial and temporal multiple counting. Since the signal pulses must be processed individually before final storage, a high processing speed is required. Even then, the application is restricted to very low pulse rates, particularly at higher pixel numbers. Indeed, the counting technique cannot remove a reduction of the DQE by signal-independent statistically occurring disturbing pulses of comparable order of magnitude (e.g., ion spots of intensifiers) [6].

1.2.3 Quantum Efficiency of Conversion Processes

The optimum design of image converter chains in accordance with the aforementioned rules requires some knowledge about the quantum efficiency of the conversion processes employed in the optoelectronic components. In particular, the interaction of the primaries with the first converting medium must be discussed.

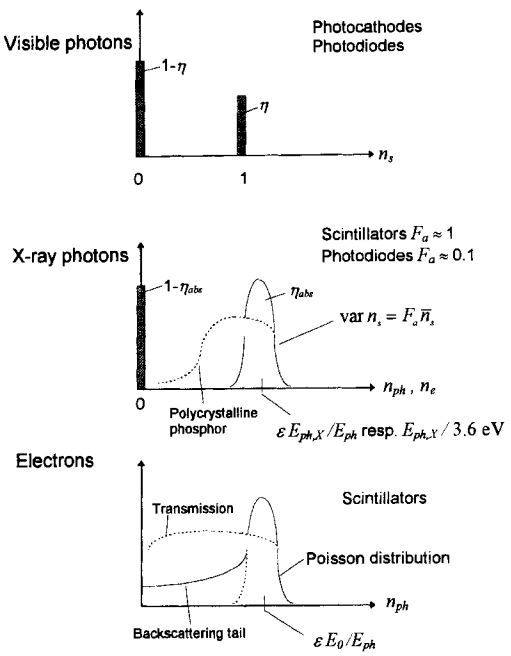


Figure 4. Types of pulse height distributions for the first converting stage using different forms of primary radiation shown schematically. F_a, Fano factor; n_s, number of secondaries.

The ultimate limit of the DQE is set by the pulse height distribution (PHD) $p_1(g_1)$ of the first stage, depending on the interaction of the different types of radiation quanta based on conversion processes to be discussed subsequently. Figure 4 provides a survey of the most important cases. Photons in the visible spectral range release only single photoelectrons with an efficiency $\bar{g}_1 := \eta$, resulting in a purely binary distribution, a model valid with photocathodes (see Sec. 1.2.3.2 of this Chapter) as well as with solid state sensors and photographic recorders (see

Sec. 1.2.3.1 of this Chapter). The PHD $p_1(g_1)$ of signals released by weak X-ray photons contains several contributions:

(a) A binary distribution of the photon energy deposits described by the absorption coefficient α and the thickness d_0 of the converting medium (see Sec. 1.5).
(b) A distribution of the number n_s of secondaries with $\text{var}\, n_s = F_a \bar{n}_s$, where the Fano factor $F_a \leqslant 1$ [8] describes a reduction of $\text{var}\, n_s$ as compared to the Poisson distribution; for solid state detectors $F_a \approx 0.1$ while in scintillators $F_a \approx 1$.
(c) A reduction of the collected signals depending on the depth of the absorption events, which is typical of attenuating scintillators such as polycrystalline phosphors (see Sec. 1.2.3.3 of this Chapter).

The detection of electrons differs from that of photons with respect to the high proportion of partial energy deposits of both back-scattered and transmitted electrons which extends the PHD by a tail to lower signals, in addition to the contributions (b) and (c) which also occur in the detection of X-rays.

Different conditions exist also with respect to the spatial resolution. In a properly designed converter chain, the shape of the PSF may be dominated by the interaction range of the primaries as well as by the dissipation of the secondaries and the procedure of their collection. In photon detectors (see Secs. 1.3 and 1.5 of this Chapter), the PSF is formed mainly by the delocalized processing of the released photoelectrons (in the solid or by an electron-optical imaging system), while in converters for fast electrons using a scintillator (see Sec. 1.4 of this Chapter) the essential contributions are caused by the electron spread in the solid and by both scattering and propagation of photons in the light-optical collection system.

The main characteristics of the most important effects will be briefly presented, with reference to the literature for greater detail.

1.2.3.1 Photographic Recording

Silver halide emulsions are still used as recorders with a high pixel number and a fairly good resolution determined by grain size and emulsion thickness. The fundamentals have been analyzed by, for example, Dainty and Show [2] and Zweig [9]. Basically, a nonlinear response occurs due to the limited number of grains each of which can detect only once. Thus, with progressing exposure the detection probability decreases. While visible photons are detected with an efficiency of the order of a few percent, energetic electrons and X-ray photons are able to hit several grains, which results in a satisfactory DQE but also in a limited dynamic range. Nonlinearity and 'fog', both dependent on the developing conditions, make the photoplate using a digitizing densitometer less suited for quantitative evaluation purposes.

1.2.3.2 Photoeffect

The release of electrons from photocathodes is widely used in photomultipliers and low-light-level image pick-up tubes [10, 11]. Numerous types of semitransparent transmission photocathodes are

available, optimized for certain wavelength ranges of the visible or near-infrared/ultraviolet spectrum and standardized using the abbreviations S1, ..., S25, the relevant data of which can be found in any brochure on multipliers, image intensifiers, or camera tubes. The advantage of photocathodes is that photoelectrons emitted into the vacuum can be accelerated in order to give significant signals of high SNR by every individual electron after subsequent conversion processes.

Using emissive material compositions, a strong photon absorption and long escape depth are realized as well as a low work function in order to extend the long-wavelength limit into the red spectrum as far as possible. At the other end of the spectrum the ultraviolet transparency of the window is an important secondary condition. The most popular photocathode, S20, is a multialkali layer (Sb–K–Na–Cs), the spectral response of which is closely matched to the sensitivity curve of the eye. At the emission wavelengths λ of some scintillators around 550 nm, corresponding to a photon energy $E_{ph} = 2.25\,\text{eV}$, these photocathodes have a sensitivity of $S_{pc} \approx 40\,\text{mA W}^{-1}$, resulting in a quantum efficiency of

$$\eta_{pc} = \frac{S_{pc} E_{ph}}{e} = \frac{S_{pc}(hc/e)}{\lambda} \approx 0.1 \qquad (12)$$

The peak sensitivity at 400 nm allows $\eta_{pc} \approx 0.25$. Special photocathodes combined with suitable window materials are available, emphasizing the ultraviolet or infrared range. The most advanced negative electron affinity cathodes using GaAs(Cs) are sensitive of up to $\lambda \approx 930\,\text{nm}$.

Since in the visible region of the electromagnetic spectrum only single electrons are released with a probability $\eta_{pc} < 1$, Eq. (11) yields a DQE $\leqslant \eta_{pc}$. Thus, a quantitative photon-counting device cannot be realized. The assumption of single electron emission is no more justified if the photon energy is increased up to the weak X-ray range when only that small part η of the impinging photons is utilized which releases photoelectrons within a small escape depth at the exit surface. Some of these may have sufficient energy to release several secondary electrons simultaneously. Hence the resulting pulse height distribution of the emitted electrons is not restricted to $g_1 = 0$ and $g_1 = 1$ but it is characterized by several peaks due to multiple events causing a further increase of the variance and thus a decrease of the DQE (see Sec. 1.5.3 of this Chapter).

The intrinsic photoeffect in semiconductors is applied in numerous television (TV) pick-up tubes of the vidicon type as well as in the modern charge-coupled devices (CCDs) (see Secs. 1.2.4.1 and 1.2.4.5 of this Chapter). Due to their low band gap, a remarkably higher yield in the formation of electron–hole pairs even in the infrared range may be obtained, which, however, can only be utilized if the noise term in Eq. (11) is suppressed by low read-out noise and/or a high storage level. The scanning electron beam read-out is not very favorable in this respect and, in addition, the storage capability of the converting layers is rather limited by recombination, but it can be very much increased by cooling. Some preferred materials are Sb_2S_3, Se, PbO, CdS, and CdSe. The most advanced technique may be the silicon photodiode employed as a mosaic target in vidicons (see Sec. 1.2.4.4) and in the sensor elements of CCDs (see Sec. 1.2.4.5 of this Chapter).

Even weak X-ray photons can be efficiently detected by collecting their photoelectrons. An impinging photon with energy $E_{ph,X}$ releases $\bar{n}_e = E_{ph,X}/E_f$ electron–hole pairs with $\text{var}\,\bar{n}_e = F_a \bar{n}_e$, where E_f is the mean formation energy and F_a is the Fano factor. Thus, for silicon, the most frequently applied semiconductor material, with $E_f = 3.6\,\text{eV}$ and $F_a \approx 0.12$ at $E_{ph,X} = 1\,\text{keV}$ a yield of 270 with a relative standard deviation of 2% can be obtained. Obviously, that fact, utilized long since in energy dispersive X-ray spectroscopy (EDX), is also very valuable in achieving a high DQE in image detectors (see Sec. 1.5.7 of this Chapter).

1.2.3.3 Scintillators

Luminescence in solids [12] has proved to be an efficient conversion process of energetic radiation (electrons and X-rays) into photons to be transferred to a light-sensitive sensor either by direct close contact or by an imaging optical element. In fact, even with X-radiation the luminescence effect in phosphors is always based on cathodo-luminescence due to the generated photoelectrons and their secondary electrons. The yield of produced photons

$$n_{ph} = \epsilon \frac{\Delta E}{E_{ph}} \tag{13}$$

is to a rough approximation determined by the energy deposit ΔE of the primary quantum, the mean energy E_{ph} of emitted photons, and an energy efficiency coefficient ϵ; further, only a fraction of the released photons can be collected due to self-absorption and the limited solid angle of acceptance of the optical element (see Sec. 1.2.3.4 of this Chapter). The energy deposit can have a broad variance for reasons which are partially somewhat different for electrons and X-rays (Fig. 4). Further, the trade-off between a high signal output on the one hand and resolution on the other must frequently be taken into account.

The retardation of primary electrons in the solid is governed by elastic and inelastic multiple scattering. A not negligible fraction of the electrons, increasing with atomic number, leaves the surface by back scattering with a wide energy distribution, after having lost a part of their initial energy by conversion processes in the solid. In addition, if a high resolution must be ensured by using scintillators that are thin compared to the penetration range of the primary electrons, the same is true for the transmitted electrons. By Monte Carlo simulation [13–15] the relevant quantities $\text{var}\,\Delta E$, $\overline{\Delta E}$, and the mean spatial density of the energy deposit within the interaction volume can readily be determined; thus, the trade-off between DQE and PSF can be balanced, as is inevitable, particularly at high electron energies.

Soft X-ray photons are converted by photoabsorption according to the absorption law with negligible scattering (see Sec. 1.5 of this Chapter). Hence the photon loses its energy $E_{ph,X}$ completely by an absorption event. Consequently, the distribution of $\Delta E \equiv E_{ph,X}$ may be much narrower, a favorable prerequisite for a high DQE. However, a more detailed consideration leads to other reasons for fluctuations of the photon yield: escape of electrons through the surfaces if the thickness becomes comparable to the range of secondary electrons, and radiation-less transitions at impurities and defects; even

the reflectivity of the surface may need to be taken into account [16].

Scintillators are commercially available as powder phosphors with grain sizes of the order of 1–10 μm, to be deposited using a binding agent on a transparent substrate and covered by a conducting aluminum mirror layer.

There are several procedures for covering a substrate: settling from suspension, cataphoretic deposition, and vapor deposition allowing a column-like orientation of CsJ crystallites with a light-guiding capability (for references, see Gruner et al. [17]). The different phosphor substances are usually characterized by a P-number [18]. If a counting technique (see Sec. 1.2.2 of this Chapter) is used, fast phosphors (e.g., P16 or P37) may be mandatory. Efficiency measurements for various phosphors with soft X-radiation are presented by Husk and Schnatterly [16] and Chappell and Murray [19].

Single crystals may be used if a restricted interaction range allows high resolution, if any fixed pattern due to the grain structure must be avoided and a narrow pulse height distribution has priority. Due to the generally relatively high refractive index n_0 of the scintillator material, however, the acceptance angle is strongly reduced by total reflection at the exit face (see Sec. 1.2.3.4 of this Chapter), while an essential part of the light propagates sideways by multiple reflections. Polycrystalline phosphors, on the other hand, offer stronger signals, however, with a higher relative variance due to the inhomogeneous intrinsic structure adding its own statistics by fluctuating interaction paths. While the light output from monocrystals (or structureless scintillator plates) has a defined total reflection limit at the emission angle $\beta_{\max} = \sin^{-1}(1/n_0)$ which is independent of the spatial distribution of the energy deposit, the output of polycrystalline phosphors is attenuated by multiple scattering and absorption, which leads to a dependence of the response on both the thickness of the screen and the localization of the primary interaction. The resulting broadening of the signal PHD impairs the DQE according to Eq. (6). Since energy deposit and photon penetration depend in an opposite sense on the thickness, the total photon output can be maximized using an optimum thickness depending on the primary energy [16].

1.2.3.4 Light Optical Elements

Only a fraction η_L of the whole photon output $\bar{n}_{ph}$ in the scintillator over the solid angle 4π as expressed by Eq. (13) can be utilized due to the limited solid angle Ω of acceptance, determined by the numerical aperture NA corresponding to a collecting angle $\beta = \sin^{-1}(\mathrm{NA}/n_0)$, and the transparency T of the chosen light optics. Thus, the light transfer element must be seen as a quantum converter stage contributing to the signal statistics by a binary distribution with a gain $\bar{g}_L := \eta_L < 1$ and $\operatorname{var} g_L = \eta_L(1 - \eta_L)$ to be introduced into Eq. (9). Generally, assuming a transparent scintillator of refractive index n_0 covered by a mirror layer of reflectivity R, the relation

$$n_L = (1+R)T\frac{\Omega}{4\pi} = (1+R)T\sin^2\frac{\beta}{2}$$
$$= \frac{1+R}{2}T\left[1-\sqrt{1-\left(\frac{\mathrm{NA}}{n_0}\right)^2}\right]$$
$$\approx \frac{1+R}{4}T\left(\frac{\mathrm{NA}}{n_0}\right)^2 \qquad (14)$$

shows the unfavorable consequences of a high refractive index n_0 as mentioned above. The available NA of the optics is mostly co-determined by the scale factor M required for a proper matching of the resolutions.

Fiber-optic plates are composed of well-ordered 6 μm fibers consisting of a core and a coat with refractive indices n_1 and n_2, respectively, which transfer the photon output of a scintillating cover layer by total reflection onto a subsequent sensor layer contacted to the plate. Large plate diameters above 50 mm and numerical apertures $NA = \sqrt{n_1^2 - n_2^2}$ of 0.66 or even 1.0 are available, resulting with $n_0 = 1.83$ (yttrium aluminum garnet, YAG) [20] in efficiencies η_L of 0.047 or 0.11, respectively. The disadvantage is the limitation of the resolution and the fixed scale factor of 1:1 (although tapers are also available) [21]. The transparency of the single fiber plate is limited to $T \approx 0.7$ due to the spatial filling factor; a sandwich of two plates, as occasionally used in coupling image intensifiers, reduces η_L by T^2 since an exact alignment of the fibers cannot be achieved. In this case, as well as with other periodic structures (e.g., CCD sensors), some Moiré effects may be observed, which can be removed together with other contributions to this 'fixed pattern' (scintillator inhomogeneity, 'chicken wire') by image processing. Most intensifier and camera tubes as well as CCDs with a larger pixel size are available with integrated fiber plates.

Lens optics allow a wide range of matching requirements by the proper choice of distances and focal lengths. The acceptance angle on the object side is in practice not completely independent of the scaling factor and the field diameter to be transferred. Some special cases may be distinguished, discussing the optics as a pair of two single lens systems both corrected for infinity but with different focal lengths f_1 and f_2 in order to realize a desired scaling factor $M = f_2/f_1$:

(1) A standard light microscope objective allows a high NA ≈ 1 with a resolution at the Abbe limit. Due to a relatively small focal length f_1 (≈ 5 mm) and the restriction of the image angle to about $\beta_i \approx 4.5°$, the object field is limited to $2r_0 = 2f_1\beta_i \approx 0.8$ mm. In order to transfer a pixel number typical for advanced image sensors, this object field limitation can only be tolerated if the high resolution of the objective can be utilized entirely. The choice of f_2 ($\gg f_1$) is then determined by the pixel size of the subsequent converter stage. The design is similar or even identical to that of a microscope with a tube length f_2.

(2) Extending the field while maintaining the NA requires a larger focal length f_1 with simultaneously increased pupil diameter, but this may not be feasible with standard light microscopic components. For this purpose a 'tandem' pair of two lenses (Fig. 5) with large apertures developed for transferring the output of X-ray intensifiers to TV pick-up tubes may be preferred [22]. The acceptance angle of such objectives is mostly characterized by their F-number. Introducing the relation $NA = (4F^2 + 1)^{-1/2}$ into Eq. (14) yields

$$\eta_L \approx (1 + R)T[4n_0^2(4F_1^2 + 1)]^{-1} \qquad (15)$$

for a single crystal with a reflecting layer.

Using, as an example, a front lens with $f_1 = 50$ mm, $F_1 = 0.75$, and $T = 0.7$ an $\eta_L = 0.032$ can be achieved as long as the whole system is magnifying ($f_2 > f_1$). If,

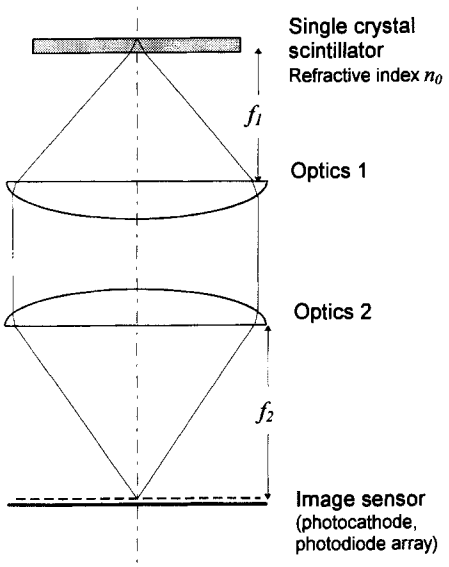


Figure 5. Schematic diagram of tandem lenses.

however, a demagnification $M = f_2/f_1 = F_2/F_1 < 1$ is required, then F_2 is limited to the same value, and we have

$$\eta_L \approx (1 + R)T \frac{M^2}{4n_0^2(4F_2^2 + M^2)} \qquad (16)$$

Thus, η_L decreases with decreasing M since, due to the limitation of f_2 and F_2, a demagnification can only be achieved by increasing f_1 and thus reducing the acceptance angle. The loss of photons occurs particularly if the whole optics are replaced by a standard photo objective.

So far the scintillator has been treated as a homogeneous refractive medium which allows an easy formulation of the optical relations. Powder phosphors for which such a treatment is not applicable may be characterized by a refractive index between 1 and 1.5 [6] and a photon absorption factor depending on the thickness of the layer. In practice, the factor $(1 + R)/n_0^2$ occurring in η_L may be determined empirically for each scintillator. Experimental experience with P20 has shown that below a mass thickness of about $5\,mg\,cm^{-2}$, absorption of photons within the scintillator can be neglected.

1.2.3.5 Secondary Emission

The emission of secondary electrons with energies below 50 eV is preferable for low-energy primary electrons. The quantitative response is well known from multipliers and scanning electron microscope instruments, and invaluable reviews are available [23, 24]. The energy dependence of the yield δ_s typically shows a limited range with $\delta_s > 1$ at a few kiloelectron volts, which is utilized for charge multiplication in photomultipliers and microchannel plates (MCPs) (see Sec. 1.2.4.3 of this Chapter). For fast primary electrons, secondary emission as the conversion process in the first stage is hardly practicable because of the adverse effect of its low yield on the DQE according to Eq. (11).

In almost all camera tubes, secondary emission is used to stabilize the potential of the target surface by the scanning electron beam releasing the video signal at the common electrode by recharging the areas discharged by the image (see Sec. 1.2.4.4 of this Chapter).

For secondary emission conduction (SEC) in transmission which was successfully utilized using thin insulating targets (KCl) with a porous structure of large internal surface area (see Sec. 1.2.4.4 of this Chapter), a yield above 10 at 7 keV was reported [25]. The advantage of such targets is their extremely long integration capability.

1.2.3.6 Electron Beam-Induced Conduction

Due to its high yield, the EBIC effect, that is, the formation of electron–hole pairs by electrons, is particularly favorable. An impinging electron with energy $E_0 = 10\,\text{keV}$ releases $\bar{n}_e \approx E_0/E_f = 2.7 \times 10^3$ electron–hole pairs in silicon ($E_f = 3.6\,\text{eV}$), the most frequently applied semiconductor material.

1.2.3.7 Imaging Plate

The imaging plate (Fig. 6) was originally developed for diagnostic X-radiography [26, 27], but it is applicable for electron radiation as well [28–31]. A storage phosphor, typically BaFBr:Eu^{2+} coated on a plastic sheet, traps the electron–hole pairs, released with high gain by the primaries, in F-centers with a high resolution determined by the range of the secondaries. This latent information is sequentially read out by a focused laser beam (630 nm) stimulating the recombination under the emission of photons (390 nm), which are detected using a photomultiplier. The merits of this device are the excellent linearity over a dynamic range of five decades and low intrinsic noise, offering the prerequisites for a high DQE. The resolution is mainly determined by scattering of the read-out beam in the layer, and depends on its thickness. The application to electrons and soft X-radiation has stimulated a reduction in the layer thickness, which was originally adapted to the absorption of hard X-rays. At present, a pixel size of 25 μm is feasible. A critical problem is the construction of highly efficient collection optics which reduce the adverse effect of scattered laser intensity [31].

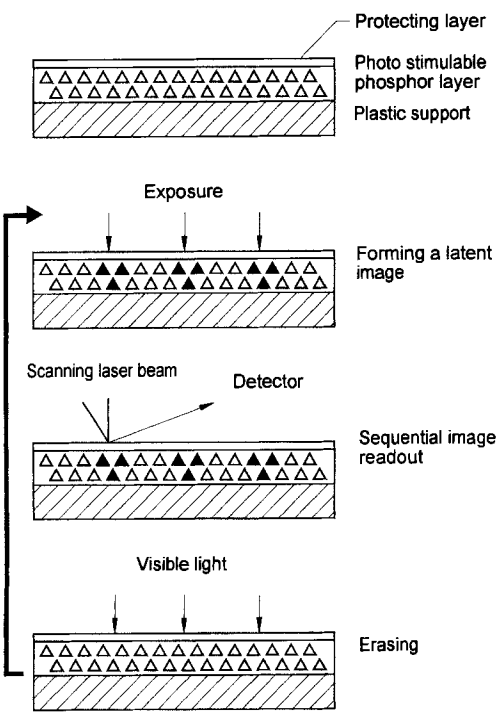


Figure 6. Working principles of the image plate.

1.2.4 Composed Systems and Optoelectronic Components

Numerous devices employ one or more of the above-mentioned physical effects in order to convert one radiation into another, to increase the quantum level, and to offer a meaningful combination for obtaining output signals suited for storage as well as for analog or digital processing. The most important components applied to on-line image recording in the various microscopes are discussed briefly in the following sections. Further details may be found in the references (e.g. [31a]).

1.2.4.1 Scintillator–Photosensor Combination

Between the emission spectrum $s(\lambda) := \Delta\bar{E}^{-1}\,dE/d\lambda$ of the scintillator and the spectral response $\eta_{pc}(\lambda)$ of the following conversion stage (e.g., a photocathode or CCD) a good overlap is required by maximizing the coupling factor

$$G := (hc)^{-1}\int s(\lambda)\eta_{pc}(\lambda)\lambda\,d\lambda \qquad (17)$$

using the data sheets of the producers (results for some combinations are given by Eberhardt [32]). Then the photoelectron number $n_e = \Delta E\,\eta_1 G$ released by an energy deposit ΔE can readily be calculated taking into account the results of Sec. 1.2.3.4 of this Chapter [nearly monochromatic emission allows the approximation $G \approx \epsilon\eta_{pc}(E_{ph})/E_{ph}$]. As an example, for the most commonly applied standard combination P20/S20, $G = 5.22\,\text{keV}^{-1}$. For the light-optical coupling the following three options are available:

(a) the direct contact to a photocathode, as used in X-ray intensifier tubes or sometimes with CCDs (see Sec. 1.5.4 of this Chapter), offers almost ideal collection efficiency;
(b) the fiber plate, as well as micro-objectives of high NA, reduces to about 1 photoelectron keV^{-1};
(c) tandem optics with $F_1 = 1.6$ allow only 0.1 photoelectron keV^{-1} [6].

The refractive index of the scintillator limits the photoelectron output. Thus the performance of the scintillator–optics–sensor combination may be characterized by the expression $\bar{n}_e/\overline{\Delta E} \approx (G/2n_0^2)\text{NA}^2$. Table 1 gives some typical figures for a selection of scintillators coupled to both a photocathode and a CCD. Obviously, due to its higher quantum efficiency the CCD is superior, particularly with scintillators at the red end of the spectrum.

These figures allow, by Eq. (11), an estimate of the DQE depending on the energy deposit which can be achieved with different radiations.

1.2.4.2 Image Intensifiers

First-generation image intensifier tubes are equipped with a transmission photocathode converting the input photon image into an electron distribution, an electrostatic electrode system accelerating the electrons by 10–20 kV onto a metallized output phosphor screen as the anode, where an amplified photon image is produced. The input and output faces are generally equipped with fiber plates, allowing the stacking of various tubes. Three designs are commercially available, which

Table 1. Data of scintillator–sensor combinations. Photoelectron yield for NA = 0.6, $T = 0.7$, and $R = 1$.

Scintillator	YAG (Ce)	CsJ (Tl)	NaJ (Tl)	CaF (Eu)	NE 102A	P20 powder
Energy efficiency, ϵ (%)	5	11.9	11.3	6.7	3.0	20
Mean photon energy, E_{ph} (eV)	2.21	2.29	2.97	2.83	2.92	2.2
Refractive index, n_0	1.83	1.80	1.85	1.44	1.58	≈1.2
Photoelectron yield, $\bar{n}_e/\Delta E$ (keV^{-1})						
with the S20 photocathode	0.074	0.02	0.26	0.26	0.094	0.75
with the TEK1024 CCD	0.35	0.73	0.29	0.30	0.11	3.41

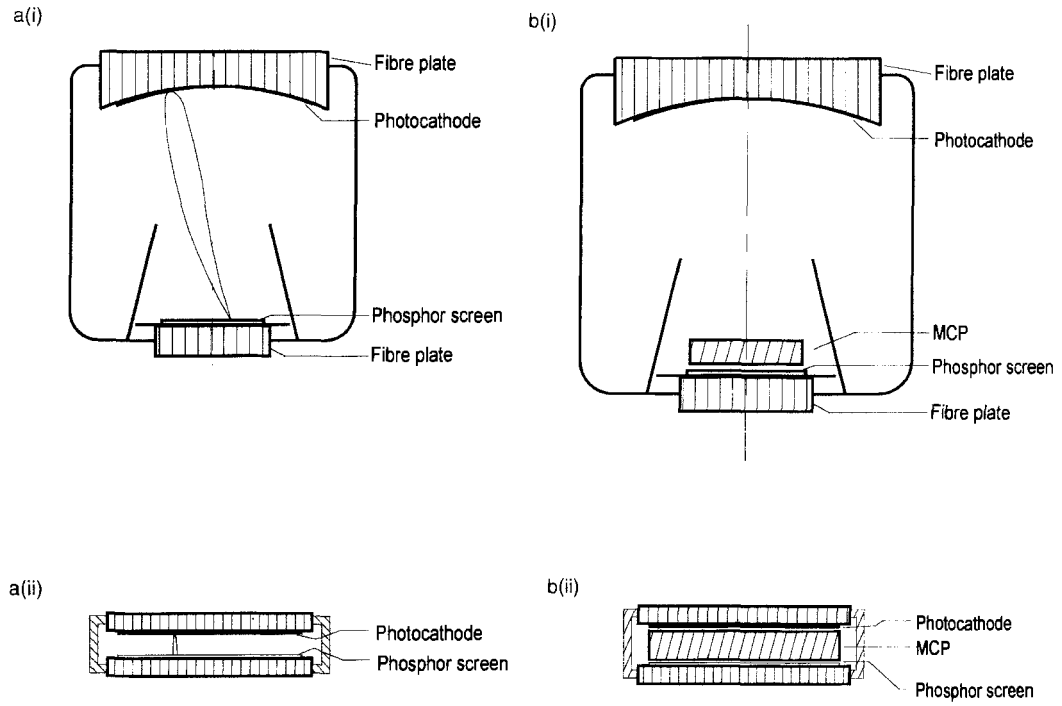


Figure 7. Examples of image intensifiers: (a) first generation and (b) second generation. (i) Electrostatically focused. (ii) Proximatively focused.

differ with respect to the focusing method (Fig. 7):

(a) magnetic focusing by embedding the tube in a longitudinal magnetic field produced by a solenoid coil;
(b) electrostatic focusing using a curved cathode and an electrical immersion lens within the tube (Fig. 7a(i));
(c) 'proximity' focusing by keeping the distance between the plane cathode and the anode as short as possible (Fig. 7a(ii)).

While (a) offers good resolution but with the drawback of large size, (b) has the advantages of more compactness, of variable gain controlled by the voltage, and of the availability of reducing systems, but some image distortion cannot be avoided. The Proxifier (c) is free from distortion, extremely compact, and can be built with a relatively large field diameter.

The gain of first-generation intensifiers can be estimated as the product of the yields of the cathode and the phosphor as

$$g \approx \eta_{\mathrm{pc}}(\lambda)\epsilon eU/E_{\mathrm{ph}} \tag{18}$$

which gives $g \approx 50$ at $\lambda = 500\,\mathrm{nm}$ if the standard S20/P20 combination and an acceleration voltage $U = 20\,\mathrm{kV}$ are used. Practically every photoelectron releases a significant pulse of $\epsilon eU/E_{\mathrm{ph}}$ photons at the output. Provided proper processing of these signals, a photon DQE $\approx \eta_{\mathrm{pc}}$ can be assumed. Intensifiers are offered with different photocathodes, allowing an adaptation of the spectral sensitivity $\eta_{\mathrm{pc}}(\lambda)$ to the

incoming spectrum as well as with different phosphors, for example with a low decay time if photon counting is required.

Second-generation intensifiers (Fig. 7b) are equipped with an MCP (see Sec. 1.2.4.3 of this Chapter) in front of the phosphor screen, the high gain of which guarantees a mean brightness level even for a very weak input. Due to their wide PHD, however, these intensifiers are only recommended for use after some preamplification or if the output pulses are electronically normalized as in single-pulse counting devices.

The resolution of the intensifiers is rather limited by the electron spread in the screen, the electron optics, the fiber plates, and, in second-generation devices, the MCP. Since the image field is limited, the transferable pixel number, which may be good enough for a standard TV technique is not sufficient for high pixel recording using advanced scientific grade CCDs (see Sec. 1.2.4.5 of this Chapter), except for Proxifiers, which are available with relatively large diameters.

1.2.4.3 Microchannel Plates

The MCP (for a review, see Lampton [33]) is a regular array of micro-tubes of internal diameter 12 μm and a length below 1 mm, the inner walls of which are covered with a semiconducting material of high secondary emission yield (Fig. 8). By a longitudinal field created by a voltage of about 1 kV across the plate, an electron released at the entrance is accelerated and multiplied by 10^4 due to repeated secondary emission. A further plate may multiply the yield once more. The high electron output can be accelerated and proximity focused at 5 keV onto a transmission phosphor

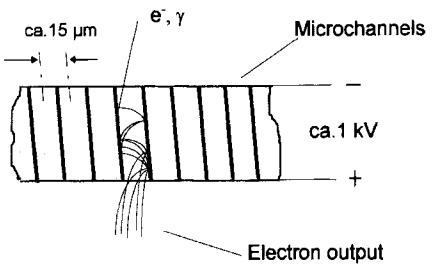


Figure 8. Principle of the MCP.

screen. The advantage of an extremely high gain is, unfortunately, accompanied by a high signal variance due to an open area ratio of 0.55, a low electron yield at the entrance for highly energetic quanta, and the fluctuation of the internal gain, which impairs the DQE according to Eq. (11) (see Sec. 1.2.2 of this Chapter) [34]. By coating the entrance faces with secondary emissive material, some improvement is possible. Moreover, the output response of high gain MCPs suffers from saturation. In pulse counting applications, this saturation effect is utilized to improve the PHD [35]; it prohibits, however, the conversion of flash images.

1.2.4.4 Television Camera Tubes

The TV technique is widely applied to convert a two-dimensional photon input into a sequential video signal in order to transfer, process, observe, and record the microscopic images on-line. Due to the TV standard, the number of pixels is limited to the order of 512^2, although the advanced high definition TV (HDTV) technique may offer some progress. Numerous types of camera tubes are available, designed for high photon sensitivity in different spectral ranges and low noise, the details of which may be found in the references [31a, 36, 37]

and, particularly, in the manufacturers' handbooks.

The basic element in a camera tube is the 'target' on which the image intensity is temporarily stored as a charge density distribution to be 'read out' periodically and linewise by a scanning electron beam of low energy which, by secondary emission, recharges the backface to cathode potential and thus induces the video signal in a common electrode for capacitive coupling out. Low noise of the subsequent video amplifier is one important prerequisite for camera performance.

The Vidicon types of tube are equipped with a semiconductor target at the entrance window, utilizing the intrinsic photoeffect for the discharge of the equilibrium potential. The selection of the material as well as its processing have been improving for many years, resulting in various special types. The variables of note for the targets are the quantum yield and the spectral response, the lateral conductivity and the storage capability depending on the dark current, the lag time, the maximum signal current I_s, and, further, the degree of avoidance of defects. The Newvicon and Chalnicon tubes, using ZnSe/ZnCdTe and CdSe targets, respectively, show a superior quantum yield over a wide spectral range, while the Saticon (SeAsTe) and Plumbicon (PbO) tubes have lower lag times at a narrower spectral response [36]. A mosaic target of silicon photodiodes constructed by advanced microlithography is an efficient way of avoiding lateral discharging. Although the quantum efficiency of the stored charge image is almost ideal, the read-out noise prevents the sensitivity from being sufficient for single-photon detection. The read-out noise from the amplifier, electron beam, and, in some cases, the target is of the order of nanoamps. Thus, for good images a relatively high exposure rate $n/\tau \approx I_s/eA$ of $>10^{10}$ photons $cm^{-2}\,s^{-1}$ is required (A is the target area of a 1 inch (2.5 cm) vidicon and τ is the frame time). Under such conditions, an excellent image quality over a dynamic ratio of some 100:1 can easily be obtained during the TV frame time.

The goal of increasing the sensitivity up to the photon noise limit requires some preamplification of about 10^5. In principle, this can be achieved using one or more intensifiers (see Sec. 1.2.4.2 of this Chapter), preferably under fiber plate coupling. More favorable, however, may be the use of a 'low-light-level' (LLL) camera tube, which are commercially available under several trade names (Fig. 9). Such tubes integrate the intensifier principle into a

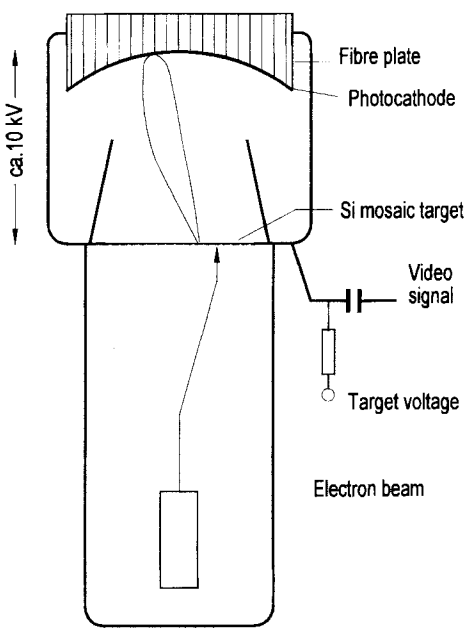


Figure 9. Schematic diagram of the 'silicon intensifier target' camera tube.

silicon target camera tube. The photoelectrons are accelerated to 10 keV and electrostatically focused onto a thinned silicon wafer having an array of diodes on the back which collects the electrons released by the EBIC effect with a yield of 2.7×10^3 (see Sec. 1.2.3.7 of this Chapter). The scanning electron beam has the same function as in vidicon tubes. These 'silicon intensifier target' (SIT) tubes [38] allow the pick-up of moonlit scenes; detecting single photoelectrons, however, requires an additional intensifier, generally coupled by a fiber plate, a compound tube frequently called an intensified SIT (ISIT).

SEC tubes [25] use a KCl target (see Sec. 1.2.3.6 of this Chapter), but are otherwise constructed like SIT tubes. Their sensitivity and dynamic range are lower; additionally, their rigidity against over-illumination is unsatisfactory. Their main merit, the storage capability, is also offered by the slow-scan CCDs, which have a far superior dynamic range.

1.2.4.5 Charge-Coupled Devices

Solid state sensors offer the most promising way for a reliable and accurate image conversion. Their enormous recent progress and continuous further improvement gives them promise as the final solution of many actual and future on-line processing tasks in microscopy. They are suitable for the recording of series as well as single images without some of the restrictions of conventional recorders. The incoming photons are accumulated on an array of silicon photodiodes to be read out once or periodically using microelectronic circuity integrated on the same chip. In earlier designs the diodes were sequentially connected by transistor switches to a common video line. A number of essential advantages, however, are offered by the CCD, in which, by MOS control, potential wells filled with electrons can be moved at high speed and with only negligible loss of charge [39–42]. By such an analogous shift register (Fig. 10a) the image information is transported in a suitable sequence to one or more common low-noise preamplifiers. Basically two read-out procedures are used (Fig. 10b):

(a) The interline transfer CCD, in which to every photodiode line a masked transport register line is attached which accepts the charge in parallel for a sequential read-out at a standard video rate.
(b) The full-frame CCD (Fig. 10c), in which the photodiodes form the transport register and are, after an exposure interval, sequentially read out with a speed adapted to the bandwidth of the noise-optimized video amplifier. The illumination must be blanked during the read-out phase unless a second interstorage area of equal format is provided (frame transfer CCD).

Procedure (a) is suited for motion picture recording at TV frequency, not only for common video cameras, but, after rapid development to higher pixel numbers, even for advanced HDTV. Motion recording in microscopy may profit from this technique. Procedure (b) is used in scientific grade CCDs, in which, by cooling, a drastic decrease of the dark current, resulting in extreme integration times, and by a slow-scan read-out, a concomitant reduction in amplifier noise is realized. These devices are highly promising in

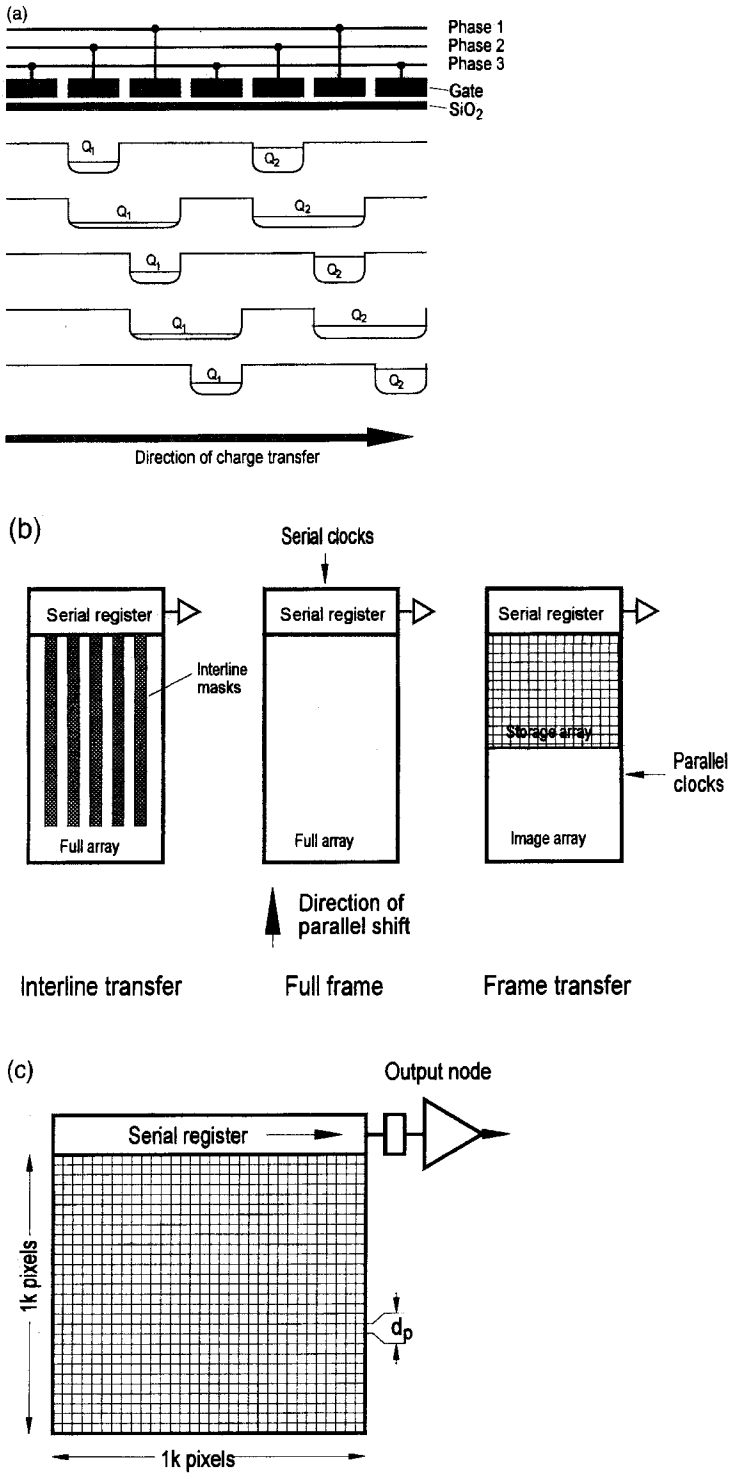


Figure 10. Charge-coupled device. (a) Principle of charge transfer in a three-phase CCD. (b) Read-out architectures. (c) Typical format of a full-frame slow-scan CCD (e.g., Table 2).

Table 2. Performance figures of the CCD Tektronix TEK 1024.

Pixel size, $d_{CCD} \times d_{CCD}$	$24 \times 24\,\mu m^2$
Pixel number	$1k \times 1k$ pixels
Full-well capacity, n_{max}	$1.5 \times 10^5\,e^-\,pixel^{-1}$
Read-out noise, n_r	
2 MHz	$<22\,e^-\,pixel^{-1}$
1.5 kHz	$<10\,e^-\,pixel^{-1}$
Dark current (−30 °C), D	$<6\,e^-\,s^{-1}\,pixel^{-1}$
Quantum efficiency (560 nm), η_{CCD}	40%
Read-out speed	
Fast	2×10^6 pixels s^{-1}
Slow	1.5×10^5 pixels s^{-1}

approaching the ideal recorder for single images.

Slow-scan CCDs (SSCCDs) have numerous merits (Table 2) [40]: a full-well capacity n_{max} above 10^5 electrons per pixel and a read-out noise n_r below 10 electrons per pixel allow an extreme dynamic range not available in other analog recorders. Negligible distortion, photometric linearity, an ideal filling factor, and independence of the signal rate up to ultrashort flash exposures are further merits. The dark current D allows satisfactory operation of video cameras at room temperature using standard frame times; in slow-scan devices, however, due to the strong temperature dependence according to $D \propto \exp[-(E_a/kT)]$ with $E_a \approx 1.05\,eV$ [41], cooling is advisable. For integration on the chip over some minutes, Peltier cooling to about −50 °C is sufficient, and by liquid nitrogen cooling a prolongation of up to hours (as required, for example, in astrophysics) is feasible. The quantum efficiency $\eta_{CCD}(\lambda)$ for electron–hole pair production by photons emphasizes the red spectral range (>0.6 at 700 nm), but it can be extended to ultraviolet wavelengths by a phosphor coating. Recently, however, back-illuminated thinned CCDs have become available where by passivation and coating techniques $\eta_{CCD} \approx 0.8$ over a wide spectral range has been achieved [41]. These will even withstand illumination with low-energy X-rays or electrons where front illumination is inapplicable due to the undesired absorption by protection and electrode layers, as well as the radiation sensitivity.

The pixel sizes vary between 8 and 27 μm with array sizes of 10^5 to above 10^7. Large pixels offer a high full-well capacity and filling factor, and better matching to some optoelectronic components, avoiding, for example, a Moiré formation between fiber plates and sensor elements. The previously mentioned low noise figures can only be achieved using a bandwidth limitation for the amplifier and analog–digital converter (ADC), resulting in a slow-scan read-out using pixel frequencies of the order of some 100 kHz (although strong progress is being made by improving the amplifiers and by parallel read out of subareas). Such sensors can be operated as quantitative image digitizers with an accuracy justifying a 14-bit ADC. Since the uniformity of the pixel response as well as of other converting elements does not have this accuracy, a numerical correction of the fixed pattern by the subsequent processing

system is advisable, using as a reference both a dark current pattern for subtraction and a uniformly illuminated image ('flat field pattern') for division.

The DQE of a converter chain using the CCD as the final element can be deduced from Eq. (7). The CCD noise consists of two statistically independent contributions expressed in the data sheets as electron numbers referred to one pixel: (a) the read-out noise as a Gaussian distribution with the standard deviation n_r, and (b) the shot noise of the dark current, which may be Poisson distributed. By adding their variances over the whole measuring area one obtains

$$DQE = \left(1 + \frac{\operatorname{var} n_e}{\bar{n}_e^2} + \frac{n_r^2 + D\tau}{\bar{N}_p \bar{n}_e^2}\right)^{-1} \qquad (19)$$

where the signal s in Eq. (7) has been replaced by the number of electrons n_e; τ is the exposure time and $\bar{N}_p$ is the number of primaries on one pixel. Considering the different primary radiations and the converting mechanisms used in front of the CCD, this fundamental relation allows the estimation and optimization of the performance, as shown below. With $(n_r^2 + D\tau)/\bar{n}_e^2 \ll 1$ a condition for single-quantum counting at low exposure is fulfilled. Already $\bar{n}_e \approx 10$ may be sufficient, which results in a saturation limit $N_{p,max} = n_{max}/\bar{n}_e \approx 10^4$ with a full-well capacity $n_{max} = 10^5$. Such an extended dynamic range is hardly achievable using other recording devices.

The development of CCD technology is making good progress, as exemplified in the *Philips Journal of Research* [42] and by Blouke [43], where a comprehensive survey on the present designs and their manufacturers is available [44]. Future developments are following several directions:

- an increase in the pixel numbers, exceeding the presently commercially offered $4k \times 4k$ CCDs;
- an increase in the read-out speed by further improving the noise figures and new organization schemes using a high degree of parallel read out;
- a higher degree of on-chip integration of control circuitry, of active control of the pixels, and of random access capability;
- the investigation of high pixel interline transfer CCDs with improved collection efficiency by means of on-chip microlenses.

1.2.5 Resolution and Sampling

The DQE as defined in Eq. (1) depends on the measuring area A, unless A is assumed to be large compared to the dimensions of the PSF. This rather trivial fact may also be expressed in Fourier space by a spatial frequency u-dependent $DQE(u) = DQE_0 \cdot MTF^2(u)$, where MTF is the modulation transfer function [2]. A comprehensive theoretical treatment of this dependence taking into account internal scattering mechanisms is given in [45, 46]. The overall MTF of the converter chain is obtained by multiplying the MTFs of the individual stages, taking into account the scaling factors between them. By proper choice of these scaling factors the stage resolutions should be matched to ensure that the response is dominated only by the first interaction processes or a subsequent sampling structure, for example, a CCD. The size d_{CCD} of the CCD pixels relative to the width of the PSF, however, should be neither too small nor too large. If it is too

small, then the fine sampling of a wide PSF leads to a coverage of only a relatively small image area by the limited number of available CCD pixels. If it is too large, then the sensor structure is fully utilized but, due to the loss of shift invariance, the common optical transfer theory becomes inapplicable [47], and the occurrence of spatial frequencies above the Nyquist frequency $(2d_{CCD})^{-1}$ may lead to artifacts in Fourier processing known as 'aliasing'. Thus, in practice, a certain bandwidth restriction is advisable by adapting the width of the PSF to one, or a few, sensor pixels.

1.3 Light Microscopy

Microphotography uses standard microscopic equipment. However, there are some special demands suggesting electronic image acquisition and processing [48]:

- visualization and recording of dynamic processes using tape or optical disks;
- contrast enhancement in real time of extremely weak image details;
- recording fluorescence images at low light levels, including radiation beyond the visible spectrum requiring some integration time and high DQE converters;
- storing and analyzing the images using computer technology.

1.3.1 Video Recording

Video cameras equipped with various types of Vidicon tubes (see Sec. 1.2.4.4 of this Chapter) or, recently, with CCDs (see Sec. 1.2.4.5 of this Chapter) are employed for real-time observation and image series registration [49]. Without any preamplifiers they work satisfactorily at higher light levels where enough electrons are accumulated on the target during the frame period to overcome the read-out noise discussed above [Eq. (7)]. Dedicated video microscopes are commercially available.

1.3.2 Low-Light-Level Detection

A considerable increase in sensitivity is obtained by SIT and ISIT tubes, such as for fluorescence microscopy [48, 50]; using MCP intensifiers even single photoelectrons at extremely low light levels may occasionally be counted. CCDs fiber-plate coupled to intensifiers are also used. Some solutions have been reviewed by Gursky and Fritz [51].

The merits of cooled SSCCDs (see Sec. 1.2.4.5 of this Chapter) are utilized for quantitative on-chip accumulation of single images with subsequent digital output [40, 52]. The DQE can be derived from Eqs. (11) and (19), separating the binary distribution and considering $\bar{n}_e = \eta_{CCD}$, with the result

$$\mathrm{DQE} = \eta_{CCD}\left(1 + \frac{n_r^2 + D\tau}{\bar{N}_p \eta_{CCD}}\right)^{-1} \qquad (20)$$

Obviously, with a pixel exposure $\bar{N}_p \gg (n_r^2 + D\tau)/\eta_{CCD}$ the DQE is determined by the efficiency $\eta_{CCD}(\lambda)$, which considerably exceeds that of photocathodes in the red spectral range and, using a fluorescent coating and/or back-illumination, in the blue and ultraviolet regions as well [41].

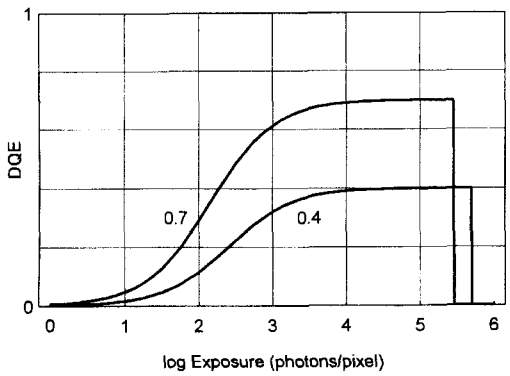


Figure 11. DQE versus exposure of a CCD for visible photons according to Eq. (20) ($\eta_{CCD} = 0.4$ and 0.7, $n_{max} = 1.5 \times 10^5\,e^-$/pixel, $n_r = 10\,e^-$/pixel $\gg D\tau$ assumed).

Consequently, the directly illuminated SSCCD allows a closer approach to the ideal detector than any photocathode device, except at low exposures (Fig. 11). Provided with digital storage and processing facilities, it may become the superior recorder in fluorescence microscopy. It has even proved successful as a sensor element in 'Nipkow disk confocal scanning microscopy' [53]. The only drawback, that is, the longer read-out time, may in time be overcome.

1.4 Electron Microscopy

The electron energy E_0 in microscopy is sufficiently high to release significant signals within an interaction volume determined by multiple scattering. The distribution $p(\Delta E)$ of energy deposits ΔE consists mainly of a peak close to E_0 with a steep slope on the high-energy side but a slow decay on the side of smaller pulse heights due to back-scattered electrons losing only part of their energy, and increases with the atomic number of the target material. With increasing energies E_0 beyond 100 keV, however, the penetration range of the electrons becomes so long that for reasons of resolution the thickness of the converting layer must be restricted. Due to the decrease in stopping power with increasing energy the mean deposit of transmitted electrons shifts to $\overline{\Delta E} < E_0$, and an additional broadening of $p(\Delta E)$ impairs the DQE. Thus, the optimization of the first converting layer with respect to resolution and the DQE is a complicated problem leading to very different results at different energies. Nevertheless, since in electron microscopes the magnification can generally be chosen within certain limits, an adaptation to a restricted resolution of the recorder is possible. Thus the image field, that is, the total number of pixels, is usually the more important parameter.

1.4.1 Photographic Recording

Direct photographic recording is the oldest and simplest read-out method (for a recent review, see Zeitler [54]). Commercial plates and films specialized for electron recording are characterized by the absence of a sensitizer and a larger layer thickness as compared to standard photoplates.

The secondary electrons released by one primary electron hit several silver halide grains. At increasing exposures the limited reserve of grains leads to an increasing nonlinearity and, finally, to saturation. The dependence of the optical density on the charge density exposure $q = en$ can be described by

$$D = D_0 + D_s[1 - \exp(-c_s q)] \qquad (21)$$

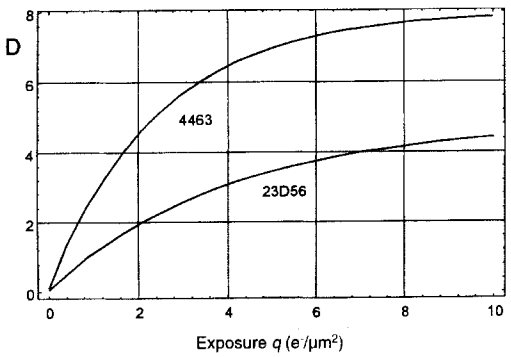


Figure 12. Density–exposure curve of Eq. (21) for two photographic emulsions (Kodak 4463, Agfa-Gevaert 23D56). (Data from Hahn [56].)

where D_0 is the 'fog', D_s is the saturation density, and c_s is a speed constant. Only with a sufficiently low exposure $q \ll c_s^{-1}$ is a linear response obtained (Fig. 12).

The DQE(q) is determined by nonlinearity, by fog, and by the statistics of the sizes and position of the grains as well as by the probability distribution for the number of grains formed by the primary. At medium electron energy it is characterized by a plateau with DQE ≈ 0.6 over about 10^2 and a decrease with lower exposures due to fog and for higher exposure due to nonlinearity according to Eqs. (7) and (3), respectively. A more detailed analysis is given by Hamilton and Marchant [55] and by Hahn [56]. In the high-energy region the DQE decreases strongly [57].

The resolution should ultimately be limited by the electron interaction volume. In practice, however, the point resolution is much worse since the saturation limit prohibits the accumulation of a number of electrons sufficient for the required SNR. As a consequence, the object detail to be resolved must be expanded to more pixels using higher magnification. Fortunately, the photoplate has a large pixel reserve. Quantitative measurements of resolution or MTF have to be processed carefully in order to correct for the electron shot noise [58]. Defining the PSF taken from such measurements as the 'true' pixel, the range of useful expositions becomes evident by plotting the DQE over the number of electrons per pixel [59]. The true pixel size is approached in low-dose microscopy of radiation-sensitive specimens where extremely noisy records of numerous identical molecular objects are superposed by the computer.

Photographic recording has the merit of simplicity, high pixel number and satisfactory DQE. Nevertheless, it does not fulfill several demands of advanced quantitative microscopy such as accuracy, reproducibility, dynamic range, and real-time response. The argument of cheapness does not hold if a costly scanning photometer for subsequent digital processing is required. Also, the decrease in cost of digital storage capacity makes electronic recording increasingly attractive.

1.4.2 Imaging Plate

The benefits of this recorder may be a DQE > 0.9 [31] and strict linearity over a dynamic ratio greater than 10^4:1, erasibility of the storage medium, easy replacement of the photoplate, and the digital output of the laser beam read-out device at 14 bits. Thus, a preferred application field is quantitative high-resolution microscopy and, in particular, diffraction, where the dynamic range of the photoplate is insufficient [28, 29]. The resolution is rather limited by the scanning laser beam

to about 25 μm, but some improvement seems to be possible [31]. The relatively high atomic number of the phosphor favors its application in high-energy microscopy [30].

1.4.3 Electronic Recording

In electron microscopy, immediate availability of the image information is wanted for three reasons:

- cinematographic recording of dynamic processes;
- alignment and focusing of the instrument at a high level of precision as a condition for subsequent photographic recording of highly resolved images;
- surpassing the photoplate in DQE, dynamic range, and processing speed for sophisticated digital image processing and low-dose techniques.

The standard fluorescent screen of the microscope allows the coarse adjustment and selection of specimen areas of interest; however, due to the short integration time of the eye, even if a binocular is used, the images, in general, cannot be observed in a way which yields information equivalent to that in a micrograph recorded on photoplate. This drawback gave the impulse to using TV chains, which after years of development, considering the points of view mentioned in Sec. 1.2.2 of this Chapter, have reached a high level of maturity, particularly by including digital frame stores and Fourier processors in order to display the power spectra. But presently the availability of high-pixel CCDs is going to remove the restriction of pixel numbers and surpass the photoplate with respect to DQE, linearity, and, in particular, to dynamic range.

1.4.3.1 Television Chains

Historically, the brightness distribution of the phosphor screen was first transferred using optics of medium speed onto a highly sensitive multistage intensifier equipped with a camera tube. This arrangement, however, seriously violated the dimensioning rules (see Sec. 1.2.2 of this Chapter). Although a gain in brightness was achieved, the observed output noise was mainly determined by the low efficiency of the optics. Then the Vidicon principle was applied in an 'open' tube attached to the microscope bottom, using the EBIC effect (see Sec. 1.2.3.6 of this Chapter) in an amorphous selenium target directly illuminated by the electrons. Unfortunately, however, the progressive recrystallization of the selenium target by the electron bombardment prohibited durable operation. Similar problems arose in the illumination of silicon mosaic targets with fast electrons. References to these and other early experimental devices are given by Herrmann and Krahl [6].

Thus, a combination is preferable which consists of a sequence P20 phosphor screen, a fiber plate, and an S20 photocathode placed within an SIT tube or an intensifier (see Sec. 1.2.4.1 of this Chapter) [6]. According to Eq. (17) and (Sec. 1.2.3.6 of this Chapter) it avoids a break of the quantum level, and one 100 kV electron produces in the target a signal above 10^5 (SIT) or even some 10^6 (ISIT). Thus, single-electron counting is possible with a DQE dominated by var ΔE. This arrangement became the standard TV converter

chain, allowing observation of images which would be invisible using the final fluorescent screen. Alternative devices are the use of a second-generation intensifier combined with a Plumbicon or Newvicon tube or, recently and now most usually, CCD video cameras.

As discussed in Sec. 1.2.1 of this Chapter, the accumulation of a sufficient SNR is required for detecting small details with low contrast. Thus, a matching of the image current density j to the storage time and, additionally, of the gain $\bar{n}_e$ to the 'dynamic window' of the target is necessary according to the balance equation $j\bar{n}_e = I_s/A$, where I_s is the signal current in the camera tube and A is the scanned target area. In TV chains this flexibility is realized only within certain limits. A current density yielding a sufficient SNR within one frame period (20 ms) is practicable at low and medium electron optical magnifications as used for video recording of dynamic processes. In this case the gain must be kept small enough to avoid target saturation, and occasionally even vidicon tubes without an intensifier may be the appropriate choice. At high magnification and with high coherence requirements, however, the current density decreases, and at the same time the gain must be increased, and the quantum noise becomes visible on the monitor. As the retention time of the eye may then become insufficient, persistent monitors are employed. The more advanced technique of accumulating sequences of TV frames, however, is the digital frame store now used which is available even on personal computers. A format of $512 \times 512 \times 8$ bit may be well matched to the capability of the TV standard. Several processing schemes are possible to realize a noise-reduced observation with a selectable persistence time, and the display of power spectra using fast Fourier processors is available. Unfortunately, every TV frame also contributes read-out noise to the image.

As a measure approaching the ideal DQE, single-electron counting (see Sec. 1.2.4.4 of this Chapter) using ISIT has been successfully applied [6, 60], which can even utilize an essential part of the back-scattering tail of the PHD, i.e. above a threshold defined by the noise of the subsequent components. The normalization procedure discriminates from low background noise as well as from fixed patterns, and the obtainable SNR is only limited by the capacity of the frame store. However, since a superposition of the spots during the lag time must be avoided this method is limited to very low current densities $j \ll 0.1\,\mathrm{pA\,cm^{-2}}$.

1.4.3.2 Slow-Scan Charge-Coupled Device Converters with a Scintillator

The demands of modern quantitative microscopy are widely met by the slow-scan CCD technique (see Sec. 1.2.4.5 of this Chapter), allowing analog image accumulation on the chip with selectable storage time and 14-bit digitization for long-term storage and numerical processing [14, 61–63]; a very comprehensive survey on the design and applications of such devices is given by de Ruijter [64]. Recently, several sophisticated schemes of information retrieval in the angstrom range, such as holography, focal series, or tilting series combination, are profiting from the merits of CCD recording, which may even be the only means for their realization. Due to the high dynamic

range, an adaptation of the gain can be dispensed with. Nevertheless, in order to utilize the advantages in an optimum manner, a quantitative analysis of the signal response over a wide range of electron energies is advisable [6, 14]. It was found that the arrangement consisting of a scintillator, light optics, and CCD offers sufficient scope to fulfill the design rules given in Sec. 1.2.2 of this Chapter. According to Eq. (13), the mean single-electron signal is $\bar{n}_e = \bar{n}_{ph}\eta_L\eta_{CCD} = \epsilon(\overline{\Delta E}/E_{ph})\eta_L\eta_{CCD}$. Considering the variance of the energy deposits ΔE and the superposed Poisson distribution of the photon numbers n_{ph} emitted at a certain ΔE, one obtains

$$\frac{\operatorname{var} n_{ph}}{\bar{n}_{ph}^2} = \frac{\operatorname{var} \Delta E}{\overline{\Delta E}^2} + \frac{1}{\bar{n}_{ph}} \tag{22}$$

and, due to the binary distributed yields of the light optics and CCD, this relation holds if n_{ph} is replaced by the single-electron signal n_e. Thus, it follows from Eq. (19) that

$$\mathrm{DQE} = \left(1 + \frac{\operatorname{var} \Delta E}{\overline{\Delta E}^2} + \frac{1}{\bar{n}_e} + \frac{1}{\bar{N}_p}\,\frac{n_r^2 + D\tau}{\bar{n}_e^2}\right)^{-1} \tag{23}$$

This relation can be straightforwardly evaluated using the data of the components if the probability distribution $p(\Delta E)$ is known. The choice of the material and thickness of the scintillator as well as that of the light optical transfer element may depend on the energy. This will be discussed briefly for single-crystalline scintillators such as YAG [20], which has proven to be a good choice with respect to the maximization of $\bar{n}_e/\overline{\Delta E}$ (see Sec. 1.2.4.1 of this Chapter and Table 1) as well as mechanical and chemical stability.

The distribution of the energy deposit $p(\Delta E)$ as well as the interaction volume, determining the resolution d_r and the required light optical scaling factor d_{CCD}/d_r, are readily accessible by Monte Carlo simulation in a layered medium [13, 14]. For an optimization, three energy ranges may be distinguished:

(a) *Low energy, 5–100 keV.* The interaction range lies below 1 μm. A light microscopic objective with NA > 0.6 allows both a satisfactory efficiency and a magnification between 10× and 40× to be chosen according to the CCD pixel size. With decreasing energy the low single-electron signal determines the DQE, and a detection of individual electrons is no longer possible. The prerequisites, design and DQE of such a device are discussed in detail by Herrmann and Sikeler [65]. Since $p(\Delta E)$ is exclusively due to backscattering, the relative variance of ΔE in Eq. (23) can also be determined using analytical models of the energy distribution of back-scattered electrons [66, 67], which is complementary to $p(\Delta E)$.

(b) *Medium energy, 50–300 keV.* The increasing penetration requires a scaling factor of the order of 1:1 for a CCD pixel size of 24 μm. Fiber plates have proven to be well suited (Fig. 13). The single-electron signal should exceed considerably the required $\bar{n}_e \approx 10$, if not, a scintillator thinner than the interaction range is used to improve the resolution. This standard converter design is commercially available and discussed in several publications [14, 62–64, 68]. By measuring the PHD using a photomultiplier with a multichannel analyzer and by Monte Carlo simulation,

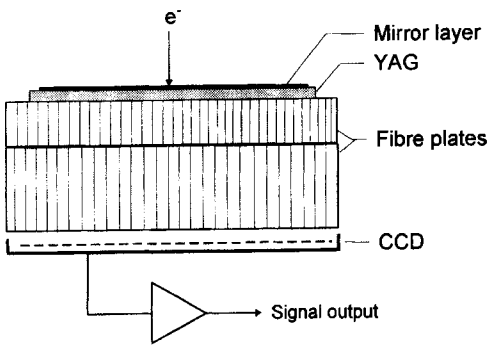


Figure 13. Schematic diagram of the fiber plate-coupled CCD detector.

a DQE of 0.9 can be verified at 100 keV, with increasing energy; however, due to increasing transmission the full absorption peak disappears, and the DQE drops to about 0.6 at 300 keV [14]. The PSF as simulated and measured by evaluating edge transient curves remains nearly independent of the energy if the YAG thickness does not exceed 30 μm [14]. Better matching of the thickness to the energy would intolerably impair the resolution as long as the 1:1 coupling by a fiber plate is used. More elaborate investigations of the shape of the PSF [69, 70] show a narrow peak placed on a relatively wide tail due to photon channeling. This tail can be removed in the processing system by deconvolution. Replacing the YAG crystal by a thinner P20 phosphor avoids this effect and allows, due to the higher photon output (see Table 1), a further improvement of resolution at the expense of DQE [68].

(c) *High energy, 400–1000 keV.* In high-energy instruments, in order to conserve resolution, considerable transmission of the electrons through the phosphor layer must be tolerated; the mean energy deposit may be estimated from the approximation

$$\overline{\Delta E} = S\rho d \tag{24}$$

where the stopping power $S := \mathrm{d}E/\mathrm{d}(\rho d)$ of the scintillator decreases with increasing energy. In order to reduce back scattering from the fiber plate as well as color center formation by the high transmitted radiation dose, a self-supporting scintillator is preferred which is slightly demagnified onto the CCD by high-speed macro-optics [71]. By Monte Carlo simulation a satisfactory DQE of 0.6 can be estimated at 1 MeV if a demagnifying tandem lens couple with $F_1/F_2 = 1.5/0.75$ is used [72]. A phosphor powder screen sedimented onto a very thin mirrored plastic foil may be preferred as the scintillator.

Table 3 summarizes some important variables for the application of these

Table 3. Data for various electron image converters using a TEK 1024 CCD ($D\tau \ll n_r^2$).

	Electron energy (keV)		
	10	100	1000
YAG scintillator thickness, d_0 (μm)	140	25	50
Optics	MO	FP	TO
Numerical aperture, NA	0.6	0.6	0.32
Single electron signal, $\bar{n}_e$ (e^-)	5	40	3
Storage capacity, N_{max} (PE pixel^{-1})	3×10^4	4×10^3	5×10^4
Maximum DQE, DQE_{max}	0.75	0.9	0.73
Lower detection limit at 0.9 DQE_{max}, N_{min}	27	0.5	73

MO, micro-optics; FP, fiber plate; TO, tandem optics.

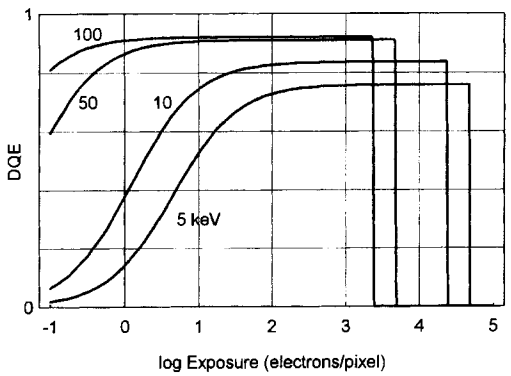


Figure 14. DQE of an electron image converter versus exposure at four different electron energies using Eq. (23). Assumptions: YAG single crystal, $d_0 >$ penetration range, NA = 1.0, $T = 0.7$, $R = 0.8$, $n_{max} = 1.5 \times 10^5\ e^-\ pixel^{-1}$, $n_r = 10\,e^-\ pixel^{-1} \gg D\tau$, and $\eta_{CCD} = 0.4$. The mean value of the energy deposits and its variance were obtained from an analytical back-scattering model [66, 67].

three devices. Figure 14 shows the evaluation of Eq. (23) using realistic parameters valid for devices (a) and (b), assuming that transmission can be neglected. The missing real-time response due to the read-out time of some seconds restricts the application of SSCCDs for immediate microscope adjustment; thus, an additional video rate camera (see Sec. 1.4.3.1 of this Chapter) is frequently used. Recently, however, a fast read-out mode ($5\,frames\,s^{-1}$) with only slightly increased noise (e.g., see Table 2) allows the combination of both functions within the same equipment [68]. Also, a rapid readout of subarrays is possible.

1.4.3.3 Directly Back-Illuminated Charge-Coupled Devices

At low energies below 5 keV the above-discussed converter chains fail due to very weak single-electron signals. While up to now the MCP (see Sec. 1.2.4.3 of this Chapter) has been the most common choice within this range, thinned CCDs may offer a promising alternative [73]. From $\bar{n}_e = \overline{\Delta E}/E_f$ with $E_f = 3.6\,eV$ a signal results which is, according to Eq. (19), sufficient by far for a high DQE even with a single-electron response. At increasing energy, however, this solution fails due to the reduction of the saturation limit $N_{p,max} = n_{max}/\bar{n}_e$.

1.5 X-Ray Microscopy

In the energy range between 100 eV and 2 keV, preferred in X-ray microscopy, the photons interact with matter mainly by photoabsorption, which exceeds scattering by a factor of 10^4. Fundamentally, the DQE in Eq. (11) is limited by the reduction factor

$$\eta_{abs} = \exp(-\alpha_d d_d)[1 - \exp(-\alpha d_0)] \qquad (25)$$

where α, α_d and d_0, d_d, respectively, are the photoabsorption coefficients and the thicknesses of the converting layer and a dead layer in front of it. Thus, a precondition for a high DQE is the complete conversion of the impinging photons into detectable photoelectrons using both $\alpha_d d_d \ll 1$ and $\alpha d_0 \gg 1$.

From the dependence of the mass absorption coefficient α/ρ on the wavelength λ and atomic number Z according to

$$\frac{\alpha}{\rho} \approx C Z^4 \lambda^3 \qquad (26)$$

(where C jumps at the characteristic edges) and using data from Henke et al. [74], certain basic tendencies can be concluded. In the low-energy region, particularly within the 'water window' below the 0_K

Table 4. X-ray absorption lengths $1/\alpha$ (μm)

$E_{ph,X}$ (eV):	100	200	500	1000	2000	5000
λ (nm):	12.39	6.19	2.48	1.24	0.62	0.25
$1/\alpha_{PMMA}$	0.4	2	1	2.5	20	
$1/\alpha_{Si}$	0.037	0.055	0.44	2.7	1.63	17.6
$1/\alpha_{Be}$	0.85	0.1	1.3	9.5	82.5	1469
$1/\alpha_{Au}$	0.03	0.046	0.043	0.104	0.44	0.70
$1/\alpha_{CsJ}$	0.03	0.31	0.55	0.25	1.17	2.55

After Henke et al. [74].

excitation energy preferred for hydrated biological samples, dead layers of even a few nanometers in thickness must be avoided. At increasing energy, however, a high conversion efficiency (e.g., due to $d_0 \geqslant 2.3/\alpha$ for $\eta_{abs} \geqslant 0.9$) offers increasing difficulties. Even if higher atomic numbers are used, the converter thickness d_0 must be restricted to maintain the resolution at the expense of the DQE. Frequently, d_0 must be chosen small enough to reduce Eq. (25) to $\eta_{abs} \approx \alpha d_0 \ll 1$. Table 4 gives some variables dependent on the wavelength for several typical materials [74]. This demonstrates the considerably different responses of detectors over the range of wavelengths used in microscopy.

The resolution is not limited primarily by the propagation of X-rays but rather by the range of Auger electrons (<5 nm) [75]. Obviously, a resolution considerably below the light microscopic limit should, in principle, be possible. In practice, however, this limit is attained only in a few devices, while others are dominated by the spread functions of subsequent signal-processing steps.

The intrinsic electron signals are generally detected by one of the following effects:

(a) chemical effects (photographic emulsion, resist);
(b) luminescence in phosphors;
(c) formation of electron–hole pairs;
(d) electron emission.

Some corresponding devices will be discussed below, considering both off- and on-line operation, as applied to contact radiography where high resolution may be required, as well as to magnifying systems where the X-ray optics may be adapted to converters with medium resolution. Integration of extremely high signal rates may be required if flash exposures using subnanosecond laser sources are applied.

1.5.1 Photographic Film and Imaging Plate

Double-sided films must be dispensed with. Special gelatine-less fine-grain photoemulsions allow a resolution of a few micrometers and a DQE of 0.1–0.3 [76].

The imaging plate (see Sec. 1.2.3.7 of this Chapter), originally developed for X-ray applications [25, 77], could be used in a thinned version as for electrons, promising a satisfactory DQE and dynamic range; however, due to the laser beam read out, the resolution may not approach the potential of soft X-radiation closely

enough. Thus, image plates appear to hold promise only for magnifying microscopes using harder radiation and for crystallography.

1.5.2 Resist

The solubility of high molecular chain molecules in organic developers, depending on the molecular weight, is increased or reduced, respectively, by X-radiation exposure due to bond scission (positive) or polymerization (negative resist) according to the absorbed energy. After developing the resist material, a relief is formed representing the spatial exposure distribution, which can be read out with sublight microscopic resolution using either transmission microscopy directly or a replica technique with shadowing, alternatively scanning electron or atomic force microscopy [78], as well as with confocal light optical microscopy [79]. Table 4 lists some absorption data, dependent on the photon energy, of polymethylmethacrylate (PMMA), the preferred resist (others are described by Valiev [80] and Seeger [81]). Obviously, in a layer 1 μm in thickness, photons are almost completely absorbed only if their energy does not exceed 500 eV. With increasing energy a considerable loss in DQE is expected; this loss can be reduced by doping the resist with heavy metals [82].

Due to the several processing steps a quantitative evaluation is hardly possible. The dissolution rate S of a polymer in a developer fluid depends on the molecular mass M according to $S \propto M^{-a}$, with $a \approx 1.5$ for PMMA developed in methyl isobutyl ketone, and thus the strongest effect is expected for high M, which is chosen to be of the order of 10^6. Since consecutive multiple scissions of one chain are possible, a strongly nonlinear characteristic curve $S(E_{abs})$ (thinning rate versus energy deposit E_{abs}, that is, the volume density of absorbed radiation energy) arises depending on the development conditions but not on the wavelength [83]. The absorption law yields

$$E_{abs}(z) = E_{in}\alpha \exp(-\alpha z) \tag{27}$$

where $E_{in} = nE_{ph}$ is the impinging local energy density. As long as $\alpha d_0 \ll 1$, the energy deposit $E_{abs} = E_{in}\alpha$ is uniformly distributed over the thickness. Thus the thickness d of the resist layer decreases linearly with the developing time t_d starting with an initial thickness d_0, that is,

$$d = d_0 - S(E_{abs})t_d = d_0 - S(E_{in}\alpha)t_d \tag{28}$$

Obviously, the achievement of a large thickness modulation $\Delta d/\Delta E_{in}$, as desired for the weak contrast images of biological samples, for example, depends on the slope dS/dE_{abs} of the characteristic curve and on the development time t_d. If the initial thickness d_0 is chosen large enough, then the weak intensity modulations can be converted to relatively large resist thickness variations compared to the remaining mean thickness. By proper matching of d_0, t_d, and E_{in} under the consideration of $S(E_{abs})$, a minimum final thickness $d_{min} > 0$ must be established. There is, however, an upper limit for the initial thickness d_0 because the resolution requirements cannot be fulfilled if d_0 is too thick, due to lateral underetching.

If, on the other hand, αd_0 is large, as required for high DQE, then the dissolution rate S decreases with progressing

mass removal, which results in a saturation of the modulation $\Delta d/\Delta E_{in}$.

X-ray resist approaches a resolution below 10 nm limited by both the molecular dimensions and the Auger electron range [75, 81]; the dynamic range, however, is poor [84], and the DQE is restricted to values around 0.1 due to the wide variance of molecule size. An estimate of the number of photons absorbed in a volume element of resolution dimensions shows that the effect of quantum noise plays a significant role [83]. Experimental evidence is found in the distinct granulation of uniformly exposed resist layers [84].

1.5.3 Transmission Photocathodes

On-line read out with a sublight microscopic resolution has been tried using a thin transmission photocathode, converting the photons into a spatial distribution of photoelectrons that are accelerated and magnified by means of a cathode lens and which impinge on a position-sensitive electron detector of medium resolution [85–88]. If the energy of the accelerated electrons exceeds about 10 keV, several types of highly efficient electron detectors are available (see Sec. 1.4.3 of this Chapter). A high DQE requires both the complete absorption of the incident photons and the efficient utilization of the photoelectron signals released in the solid. Following a simple model assumption [89, 90], the electron yield η_e can be estimated by integrating over the z-coordinate the product of the photon absorption probability, given by Eq. (27), and the escape probability, again an exponential containing the mean free path λ_e of the electrons, with the result

$$\eta_e = \int_0^{d_0} \alpha \exp(-\alpha z) \exp\left(-\frac{d_0 - z}{\lambda_e}\right) dz$$

$$= \frac{\lambda_e \alpha}{1 - \lambda_e \alpha} \left[\exp(-d_0 \alpha) - \exp\left(-\frac{d_0}{\lambda_e}\right)\right] \tag{29}$$

As a consequence, there exists an optimum thickness $d_{opt} = \lambda_e(\lambda_e\alpha - 1)^{-1} \ln(\lambda_e\alpha)$ of the photocathode for which η_e attains a maximum value $\eta_{max} = (\lambda_e\alpha)^{1/(1-\lambda_e\alpha)}$ (Fig. 15). Thus, materials are required offering a large value $\lambda_e\alpha$. Isolators, for example CsJ, may be superior to metals [89, 91]. Unfortunately, the DQE cannot be identified with the electron yield η_e since the electron emission does not follow a binary distribution as for visible radiation [90]. The energy of the X-ray photons is sufficient to release multiple electrons in each absorption event, and thus the

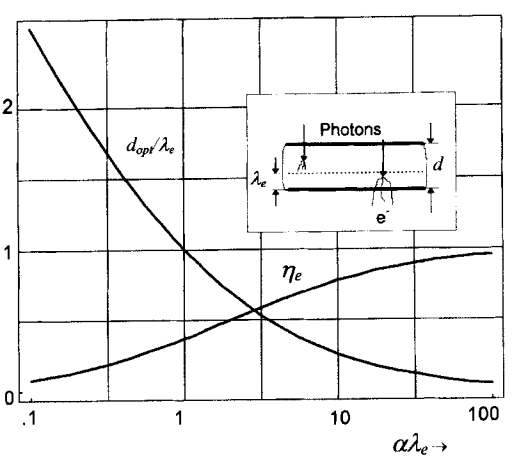


Figure 15. X-ray transmission photocathodes. Optimum relative thickness d_{opt}/λ_e and corresponding electron yield versus the normalized absorption coefficient $\lambda_e\alpha$ (λ_e is the mean free path length of secondary electrons in the solid).

emitted signals show an additional PHD, which, according to Eq. (11), leads to a further impairment of the DQE against η which must now be defined as the probability of releasing a count independent of its weight. Estimates and measurements of the DQE using gold foils [92] indicated values of the order of a few percent with Ti K_α radiation. An increase of the DQE with increasing wavelength is expected.

The resolution of the transmission photocathode, on the other hand, is limited to the order of the mean free path length λ_e, which lies in the region of 4 nm (Au) and 25 nm (CsJ). The subsequent electron–optical processing system must be well designed in order to maintain this resolution. A cathode lens, as customary in various types of electron emission microscopes, accelerates the electrons to about 20 keV, and an additional lens inserted into the anode section produces a focused image at a magnification of about 20. The resolution of cathode lenses is dominated by the chromatic aberration, which depends to a rough approximation [93] on the energy width ΔE_{pe} of the emitted electrons and the electric field strength F_e at the cathode surface according to

$$\delta \approx \Delta E_{pe}/F_e \tag{30}$$

For a more detailed analysis considering the other aberrations, see Polack and Lowenthal [86]. While for technical reasons the field strength is limited to about 100 kV cm^{-1}, the energy width may be considerably higher than in typical photoemission microscopes. There are two solutions: a diaphragm placed in the diffraction plane of the cathode lens and/or an imaging energy filter allowing a selectable restriction of the energy spectrum contributing to the image. The Castaing–Henry system using an electrostatic mirror combined with a magnetic 90° deflector [94] as well as the purely magnetic Ω-filter [95] are high-performance solutions to this problem, but unfortunately not achievable without some effort. In addition, the filtering may remove some signal pulses and thus further reduce the DQE.

1.5.4 Microchannel Plates

Due to its compactness the MCP is frequently used in X-ray astronomy combined with resistive or structurized anodes as position sensitive photon-counting detectors (review by Fraser [96]). The quantum efficiency for soft X-rays is between 1 and 10%; however, by depositing a material with a high secondary electron yield (e.g., MgF_2 or CsJ) an improvement of up to 50% can be achieved [97]. The high electron gain allows image pick-up from a subsequent phosphor screen by standard video cameras. The wide PHD, however, reduces the DQE and makes a spectroscopic resolution impossible. The saturation prohibits an application to flash exposures.

1.5.5 Television Chains

The design and application of TV pick-up devices for image display in real time has been reviewed [98, 99]. The rules of their design have to consider the general relations given in Sec. 1.2 of this Chapter requiring, in particular, a highly efficient photon conversion in the entrance stage by both a high photon absorption and access

to the secondary signals which is as complete as possible. While a few attempts with X-ray-sensitive pick-up tubes have successfully been made at higher photon energies, most devices for soft radiation utilize the window-less scintillators and standard optoelectronic components (see Sec. 1.2.4 of this Chapter) of LLL camera technology.

1.5.5.1 X-ray-Sensitive Vidicons

Several designs of X-ray-sensitive Vidicons have been reported (reviewed by Hartmann [99]) in which a beryllium window and a photoconductive target optimized for X-rays is generally used. These special tubes were developed for live topography using energies above a few kiloelectronvolts, and may, due to the beryllium window (see Table 4), be completely unable to detect very soft X-radiation with a high efficiency. In order to increase the X-ray absorption, a PbO target 15 μm in thickness in a Vidicon tube has been used; with Cu K_α radiation (8 keV) a DQE of 78% and a resolution below 25 μm were reported [100].

1.5.5.2 Conversion to Visible Radiation

The basic physical principle of the classical X-ray intensifier is the conversion of the photons to electrons using a phosphor layer in close contact with a photocathode, the emitted electrons from which are accelerated and demagnified onto a highly resolving screen. Although this device has sometimes been used for topography, it is not applicable to microscopy, for the reasons mentioned above. For microscopy a window-less conversion using an open phosphor is the best choice. The application of phosphors with subsequent LLL processing does not differ much from the devices mentioned above (see Sec. 1.2.4.4 of this Chapter). Generally the visible photons are fed using a fiber plate to the photocathode of a SIT tube or an intensifier (single or multiple stage) coupled to a Vidicon or a CCD. The resolution is limited by the light spread in the phosphor layer, depending on the thickness, and by the fiber plate structure. Phosphors with a high stopping power are preferred in order to optimize the resolution and photon output. Unfortunately, the photon attenuation in the phosphor prohibits full absorption with hard X-rays. This can be addressed using Eq. (29) by identifying λ_e with an empiric photon attenuation length λ_{ph} [16], with the conclusion that an optimum thickness of the phosphor screen exists offering an efficiency considerably below unity if $\alpha\lambda_{ph} \ll 1$.

Again the number of photoelectrons $\bar{n}_e$ released by one absorbed X-ray photon of energy $E_{ph,X}$ ($\equiv\overline{\Delta E}$) can be obtained from Sec. 1.2.4.1 of this Chapter and Table 1. Obviously, only photons above 1 keV can release photoelectrons with an acceptable yield. Their multiplication in order to produce a charge spot on the target strong enough to overcome the read-out noise of the camera follows the general rules already treated in Sec. 1.2.4.4 of this Chapter.

1.5.6 Slow-Scan Charge-Coupled Device Chains

Cooled slow-scan CCDs may allow the construction of the most advanced X-ray

Table 5. Data for an X-ray image converter using a TEK 1024 CDD ($D\tau \ll n_r^2$). Assumptions: CsJ(Tl) crystal, NA = 1.0, $T = 0.7$, $R = 1$ (= 0 for $E_{ph,X} \leqslant 500\,eV$), $N_{p,min}$ for 0.9 DQE_{max}, $d_{0,min}$ for $\eta_{abs} = 0.99$.

$E_{ph,X}$ (eV):	100	200	500	1000	2000	5000
λ (nm):	12.39	6.19	2.48	1.24	0.62	0.25
$\bar{n}_e$ (e^-)	0.12	0.24	0.6	2.03	4.06	10.1
DQE_{max}	0.11	0.2	0.39	0.67	0.8	0.91
$N_{p,max}$ (photon pixel^{-1})	12×10^5	6×10^5	2.4×10^5	7.5×10^4	3.8×10^4	1.5×10^4
$N_{p,min}$ (photon pixel^{-1})	6660	2990	921	146	43	8
$d_{0,min}$ (μm)	0.15	1.4	2.5	1.15	5.4	11.7

image read-out device, characterized by quantum-limited recording, extreme storage capability, independency of the signal rate, excellent linearity, pixel numbers comparable to photographic recording, and, due to the 14-bit digital output, online transfer to high-performance image-processing systems. Using an open scintillator avoids any photon losses by a window, and a versatile choice of material and thickness of the scintillator is possible. Combining Eqs. (11), (19), and (25), the DQE can be predicted to be

$$DQE = \eta_{abs}\left(1 + \frac{1}{\bar{n}_e} + \frac{n_r^2 + D\tau}{\eta_{abs}\bar{N}_p\bar{n}_e^2}\right)^{-1} \quad (31)$$

The variance contribution of the energy deposit was neglected in this case because a partial absorption of the photon energy may not be expected.

Table 5 shows the evaluation of Eq. (31) at various photon energies assuming, as an example, a mirrored CsJ scintillator (see Table 1), optics with NA = 1.0, and a TEK 1024 CCD (see Table 2). Obviously, a satisfactory DQE is feasible above 500 eV only by using some signal integration. In order to keep the reduction factor η_{abs} as close as possible to unity a scintillator thickness $d_0 \geqslant 2.3/\alpha$ may be chosen which sets different resolution limits over the energy range and consequently requires different scale factors for the optics. At higher energy the DQE decreases since full absorption cannot be maintained. At low energy the decreasing single-electron signal dominates the response and, additionally, the gain by the aluminum mirror layer is overcompensated by the X-ray absorption according to Eq. (25). Thus, at very low photon energies, a mirror-less scintillator is superior. Using single-crystal scintillators, even the light microscopic resolution limit may be attained (see Sec. 1.2.3.4 of this Chapter).

1.5.7 Directly Illuminated Charge-Coupled Device Sensors

In principle, a soft X-ray photon produces a significant number $\bar{n}_e = E_{ph}/3.6\,eV$ of electron–hole pairs with $\mathrm{var}\, n_e = F_a\bar{n}_e$ ($F_a \approx 0.12$ Fano factor) in silicon (see Sec. 1.2.3.2). If recombination is avoided by the collecting field in the diodes this signal considerably exceeds the read-out noise contributions and allows excellent spectroscopic resolution as well as a single-quantum response over the whole range of photon energies E_{ph} relevant for X-ray microscopy. Even spatially resolved energy-dispersive spectroscopy is possible [101] as long as an overlap of signals on one pixel is avoided by low exposure and

the read-out signal is fed via a multichannel analyzer to a subsequent multiple frame storage system. On the other hand, the

$$DQE_{max} = \eta_{abs}(1 + F_a\, 3.6\,eV/E_{ph})^{-1} \approx \eta_{abs} \quad (32)$$

is rather dominated by the probability of forming a signal within the wells, as given by their geometrical dimensions and the absorption coefficient of silicon according to Eq. (25). Unfortunately, with increasing energy the saturation limit n_{max} of the CCD limits drastically the $SNR \leqslant \sqrt{\eta_{abs} N_{p,max}} = n_{max}\, 3.6\,eV/E_{ph}$.

In practice some physical or technical impediments must be considered. Commercial CCDs are covered by a protective Si_3N_4 layer about 2 μm in thickness, which is transparent to visible light but strongly absorbs soft X-rays. Together with the metallic control electrodes, a dead layer is formed for soft radiation. Harder X-rays, on the other hand, are less absorbed in the dead layer, but also less so in the depletion zone. They may additionally impair the function of the MOS read-out elements by radiation damage [102]. Two possible remedies are: (a) covering the protected input face by a phosphor layer that converts each absorbed X-ray photon into a number of visible photons [103, 104], and (b) using thinned and back-illuminated CCDs [41, 105, 106] (see Sec. 1.2.4.5).

By using highly resistive material the depletion zone can be far extended, but its close approach to the back face of the thinned CCD offers a hard technological problem. A field-free charge-trapping dead layer remains. Using realistic values of $d_d/d_0 = 50\,nm/20\,\mu m$ [106] and the absorption coefficient of silicon (see Table 4), from Eq. (25), results in $DQE \approx 0.3/0.98/0.68$ at photon energies $E_{X,ph} = 0.1/1/5\,keV$ (Fig. 16). Thus, using remedy (b) the dead layer problem at low energies is not entirely removed. Method (a) avoids a dead layer and allows the use of higher absorbing phosphors as well as components of consumer video cameras, while remedy (b) requires a more costly special design of a scientific-grade CCD and electronics. In any case, the resolution is limited by the sensor element size to about 10 μm at best.

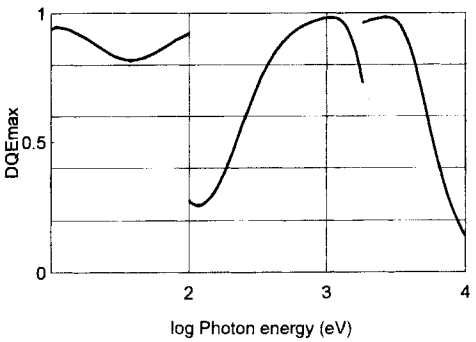


Figure 16. DQE_{max} of a directly X-ray back-illuminated CCD versus photon energy using Eqs. (25) and (32) and silicon absorption data from Henke et al. [74]. Assumptions: depletion depth $d_0 = 20\,\mu m$, dead layer thickness $d_d = 50\,nm$, and Fano factor $F_a = 0.12$.

Acknowledgment

I thank my colleague F. Lenz for critically reading the manuscript and for many valuable suggestions.

1.6 References

[1] A. Rose, *J. Soc. Mot. Picture Eng.* **1946**, *47*, 273.

[2] J. C. Dainty, R. Show, *Image Science*, Academic Press, London **1974**.

[3] C. Jones, *J. Opt. Soc. Am.* **1959,** *49*, 645.
[4] L. Mandel, *Br. J. Appl. Phys.* **1959,** *10*, 233.
[5] E. Breitenberger, *Progr. Nucl. Phys.* **1955,** *4*, 56.
[6] K.-H. Herrmann, D. Krahl, *Adv. Opt. El. Micr.* **1984,** *9*, 1.
[7] A. Boksenberg, C. I. Coleman, *Adv. Electron. Electron Phys.* **1979,** *52*, 355.
[8] U. Fano, *Phys. Rev.* **1947,** *72*, 26.
[9] H. J. Zweig, *J. Opt. Soc. Am.* **1961,** *51*, 310.
[10] P. N. J. Dennis, *Photodetectors: An Introduction to Current Technology*, Plenum Press, New York **1985.**
[11] M. Jedlicka, P. Kulhánek, *Vacuum* **1986,** *36*, 515.
[12] H. P. Kallmann, G. M. Spruch, *Luminescence of Organic and Inorganic Materials*, Wiley, New York **1962.**
[13] D. C. Joy, *Inst. Phys. Conf. Ser., EUREM 88, York* **1988,** *93/I*, 23.
[14] I. Daberkow, K.-H. Herrmann, L. Libin, W.-D. Rau, *Ultramicroscopy* **1991,** *38*, 215.
[15] M. Kotera, Y. Kamiya, *Ultramicroscopy* **1994,** *54*, 293.
[16] D. E. Husk, S. E. Schnatterly, *J. Opt. Soc. Am. B.* **1992,** *9*, 660.
[17] S. M. Gruner, J. R. Milch, G. T. Reynolds, *Nucl. Instrum. Methods* **1982,** *195*, 287.
[18] JEDEC, *Optical Characteristics of Cathode Ray Tube Screens*, Electrical Industry Association, Washington DC. **1975.**
[19] J. H. Chappell, S. S. Murray, *Nucl. Instrum. Methods* **1984,** *221*, 159.
[20] R. Autrata, P. Schauer, Jiri Kvapil, Josef Kvapil, *Scanning* **1983,** *5*, 91.
[21] Galileo Electro-Optics Corp. *Fibre Optics: Theory and Application, Technical Memorandum 100, Fibre Optic Faceplates, Data Sheet 1000.* Gallileo Electro-Optics Corp., Sturbridge, MA.
[22] Optische Werke G. Rodenstock, *X-Ray Manual*, Optische Werke G. Rodenstock, München **1992.**
[23] L. Reimer, *Scanning Electron Microscopy, Springer Series in Optical Science*, Vol. 45, Springer, Berlin **1985.**
[24] L. Reimer, *Image Formation in Low-Voltage Scanning Microscopy, Tutorial Texts in Optical Engineering*, No. 12, SPIE Optical Engineering Press, Bellingham **1993.**
[25] G. W. Goetze, *Adv. Electron. Electron Phys.* **1966,** *22A*, 219.
[26] M. Sonoda, M. Takano, J. Miyahara, H. Kato, *Radiology* **1983,** *148*, 833.
[27] Y. Amemiya, K. Wakabayashi, H. Tanaka, Y. Ueno, J. Miyahara, *Science* **1987,** *237*, 164.
[28] N. Mori, T. Oikawa, T. Katoh, J. Miyahara, Y. Harada, *Ultramicroscopy* **1988,** *25*, 195.
[29] D. Shindo, K. Hiraga, T. Oikawa, N. Mori, *J. Electron Microsc.* **1990,** *39*, 449.
[30] S. Isoda, K. Saitoh, T. Ogawa, S. Moriguchi, T. Kobayashi, *Ultramicroscopy* **1992,** *41*, 99.
[31] C. Burmester, H. G. Braun, R. R. Schröder, *Ultramicroscopy* **1994,** *55*, 55.
[31a] L. M. Biberman, S. Nudelman (Eds.), *Photoelectronic Imaging Devices, Vol. I and II*, Plenum Press, New York **1971.**
[32] E. H. Eberhardt, *Appl. Optics* **1968,** *7*, 2037.
[33] M. Lampton, *Sci. Am.* **1981,** *245(5)*, 46.
[34] K.-H. Herrmann, M. Korn, *J. Phys. E: Sci. Instrum.* **1987,** *20*, 177.
[35] Galileo Electro-Optics Corp. *Microchannel Plates, Data Sheet 9000, Detector Assemblies, Data Sheet 7000.* Galileo Electro-Optics Corp., Sturbridge, MA.
[36] R. G. Neuhauser, *J. SMPTE* **1987,** *96*, 473.
[37] B. L. Morgan (Ed.), *Adv. Electron. Electron Phys.* **1988,** *74.*
[38] G. A. Robinson, *J. SMPTE* **1980,** *89*, 249.
[39] D. F. Barbe (Ed.), *Charge Coupled Devices, Springer Topics in Applied Physics*, Vol. 38, Springer, Berlin **1980.**
[40] Photometrics Ltd., *Charge-Coupled Devices for Quantitative Electronic Imaging*, Photometrics Ltd., Tucson **1989.**
[41] R. P. Khosla, *Microelectr. Eng.* **1992,** *19*, 615.
[42] *Philips. J. Res.* [Special issue] **1994,** *48*, 3.
[43] M. L. Blouke (Ed.), *SPIE Proc.* **1994,** *2172.*
[44] E. R. Fossum, *SPIE Proc.* **1994,** *2172*, 38.
[45] M. Rabbani, R. van Metter, *J. Opt. Soc. Am. A* **1989,** *6*, 1156.
[46] H. Mulder, *J. Opt. Soc. Am. A* **1993,** *10*, 2038.
[47] J. C. Feltz, M. A. Karim, *Appl. Opt.* **1990,** *29*, 717.
[48] D. Shotton (Ed.), *Techniques in Modern Biomedical Microscopy*, Wiley-Liss, New York **1993.**
[49] S. Inoué (Ed.), *Video Microscopy*, Plenum Press, New York **1986.**
[50] T. T. Tsay, R. Inman, B. Wray, B. Herman, K. Jacobsen, *J. Microsc.* **1990,** *160*, 141.
[51] H. Gursky, G. Fritz, *Scanning* **1991,** *13*, 41.
[52] R. K. Aikens, D. A. Agard, J. W. Sedat, *Methods Cell Biol.* **1989,** *29*, 291.
[53] B. R. Masters, G. S. Kino, *Electronic Light Microscopy*, Wiley-Liss, New York **1993,** p. 315.
[54] E. Zeitler, *Ultramicroscopy* **1992,** *46*, 405.
[55] J. F. Hamilton, J. C. Marchant, *J. Opt. Soc. Am.* **1967,** *57*, 232.
[56] M. Hahn, *Electron Microscopy at Molecular Dimensions* (Eds.: W. Baumeister, W. Vogell) Springer, Berlin **1980,** p. 200.
[57] D. G. Ast, *J. Appl. Phys.* **1974,** *45*, 4638.
[58] D. A. Grano, K. H. Downing, *38th Annu. EMSA Meeting* **1980,** 228.

[59] K.-H. Herrmann, *Electron Microsc.* **1982, 1, 131.**
[60] K.-H. Herrmann, D. Krahl, H.-P. Rust, O. Ulrichs, *Ultramicroscopy* **1978,** *3*, 227.
[61] J. C. H. Spence, J. M. Zuo, *Rev. Sci. Instrum.* **1988,** *59*, 2102.
[62] S. Kujawa, D. Krahl, *Ultramicroscopy* **1992,** *46*, 395.
[63] O. L. Krivanek, P. E. Mooney, *Ultramicroscopy* **1993,** *49*, 95.
[64] W. J. de Ruijter, *Micron* **1995,** *26*, 247.
[65] K.-H. Herrmann, R. Sikeler, *Optik* **1995,** *98*, 119.
[66] H. Niedrig, *Electron Beam Interactions with Solids for Microscopy, Microanalysis & Microlithography* (Eds. D. F. Kyser, H. Niedrig, D. E. Newbury, R. Shimizu), Scanning Electron Microscopy, Inc., AMF O'Hare **1982,** p. 51..
[67] U. Werner, H. Bethge, J. Heydenreich, *Ultramicroscopy* **1982,** *8*, 417.
[68] I. Daberkow, K.-H. Herrmann, L. Liu, W. D. Rau, H. Tietz, *Ultramicroscopy*, submitted.
[69] A. L. Weickenmeier, W. Nüchter, J. Mayer, *Optik* **1995,** *99*, 147.
[70] J. M. Zuo, *Electron Microsc.* **1994,** *1*, 215.
[71] G. Y. Fan, M. H. Ellisman, *Ultramicroscopy* **1993,** *52*, 21.
[72] K.-H. Herrmann, L. Liu, *Optik* **1992,** *92*, 48.
[73] D. G. Stearns, J. D. Wiedwald, *Rev. Sci. Instrum.* **1989,** *60*, 1095.
[74] B. L. Henke, P. Lee, T. J. Tanaka, R. L. Shimabokuro, B. K. Fujikawa, *Atomic Data Nucl. Data Tables* **1982,** *27*, 1.
[75] K. Murata, M. Tanaka, K. Kawata, *Optik* **1990,** *84*, 163.
[76] L. Jochum, *SPIE Proc.* **1986,** *733*, 492.
[77] J. Miyahara, K. Takahashi, Y. Amemiya, N. Kaniya, Y. Satow, *Nucl. Instrum. Methods A* **1986,** *246*, 572.
[78] T. Tomie, H. Shimizu, T. Majima, M. Yamoda, T. Kanayama, H. Kondo, M. Yano, M. Ono, *Science* **1991,** *252*, 691.
[79] P. C. Cheng, D. M. Shinozaki, *X-Ray Microscopy III* (Eds.: A. G. Michette, G. R. Morrison, C. J. Buckley), Springer, Berlin **1992,** p. 335.
[80] K. Valiev, *The Physics of Submicron Lithography*, Plenum Press, New York **1992.**
[81] D. Seeger, *IBM J. Res. Dev.* **1993,** *37*, 435.
[82] I. Haller, R. Feder, M. Hatzakis, E. Spiller, *J. Electrochem. Soc.* **1997,** *126*, 154.
[83] E. Spiller, R. Feder, *Topics Appl. Phys.* **1977,** *22*, 35.
[84] D. M. Shinozaki, *X-Ray Microscopy II* (Eds.: D. Sayre, M. Howells, J. Kirz, H. Rarback), *Springer Series in Optical Sciences*, Vol. 56, Springer, Berlin **1987,** p. 118.
[85] L. Y. Huang, *Z. Phys.* **1957,** *149*, 225.
[86] F. Polack, S. Lowenthal, *X-Ray Microscopy* (Eds.: G. Schmahl, D. Rudolph), *Springer Series in Optical Sciences*, Vol. 43, Springer, Berlin **1984,** p. 251.
[87] F. Polack, S. Lowenthal, D. Phalippou, P. Fournet, *X-Ray Microscopy II* (Eds.: D. Sayre, M. Howells, J. Kirz, H. Rarback), *Springer Series in Optical Sciences*, Vol. 56, Springer, Berlin **1987,** p. 220.
[88] K. Kinoshita, T. Matsumura, Y. Inagaki, N. Hirai, M. Sugiyama, H. Kihara, N. Watanabe, Y. Shimanuki, A. Yagashita, *X-Ray Microscopy III* (Eds.: A. G. Michette, G. R. Morrison, C. J. Buckley), *Springer Series in Optical Sciences*, Vol. 67, Springer, Berlin **1992,** p. 335.
[89] B. L. Henke, J. P. Knauer, K. Premaratne, *J. Appl. Phys.* **1981,** *52*, 1509.
[90] G. W. Fraser, *Nucl. Instrum. Methods A* **1985,** *228*, 532.
[91] A. Akkerman, A. Gibrekhterman, A. Breskin, R. Chechik, *J. Appl. Phys.* **1992,** *72*, 5429.
[92] W. Mullion, Diploma Thesis, Tübingen University **1993.**
[93] A. Recknagel, *Z. Phys.* **1941,** *117*, 689.
[94] R. Castaing, L. Henry, *C. R. Acad. Sci. Paris* **1963,** *B255*, 76.
[95] S. Lanio, H. Rose, D. Krahl, *Optik* **1986,** *73*, 56.
[96] G. W. Fraser, *Nucl. Instrum. Methods* **1984,** *221*, 115.
[97] G. W. Fraser, *Nucl. Instrum. Methods* **1983,** *206*, 251, 265.
[98] S. M. Gruner, *Rev. Sci. Instrum.* **1989,** *60*, 1545.
[99] W. Hartmann, *Topics Appl. Phys.* **1977,** *22*, 191.
[100] R. Nishida, S. Okamoto, *Rep. Res. Inst. Electron., Shizuoka University* **1966,** *1*, 21, 185.
[101] D. H. Lump, G. R. Hopkinson, A. Wells, *Nucl. Instrum. Methods A* **1984,** *221*, 150.
[102] N. M. Allinson, D. W. E. Allsop, J. A. Quayle, B. G. Magorrian, *Nucl. Instrum. Methods A* **1991,** *310*, 267.
[103] P. Burstein, J. M. Davis, *X-Ray Microscopy* (Eds.: G. Schmahl, D. Rudolph), *Springer Series in Optical Sciences*, Vol. 43, Springer, Berlin **1984,** p. 184.
[104] R. Germer, W. Meyer-Ilse, *Rev. Sci. Instrum.* **1986,** *57*, 426.
[105] C. Castelli, A. Wells, K. McCarthy, A. Holland, *Nucl. Instrum. Methods A* **1991,** *310*, 240.
[106] J. L. Culhane, *Nucl. Instrum. Methods A* **1991,** *310*, 1.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

2 Image Processing

2.1 Introduction

Image processing is a generic term which covers several activities related to digital images, that is, images which have been converted into arrays of pixels, as described in Sec. 1 of this Chapter. These topics concern image preprocessing, image restoration, image segmentation, image analysis, three-dimensional reconstruction, etc.

The microscopes described in the previous chapters are able to produce different kinds of image. Although the simple observation of these images is often sufficient for interpreting the specimen, it is increasingly necessary to manipulate the images, in order either to help the qualitative interpretation or to extract some quantitative information.

Unfortunately, it is not so easy to go from image acquisition to data exploitation through image analysis. Usually, a number of intermediate steps must be devised (Fig. 1). Image preprocessing consists of preparing images for subsequent operations. Image processing (in the restricted sense of the term) consists of transforming an image into another one, which is intended to be more manageable, either in terms of interpretation or in terms of subsequent analysis. Image analysis consists of transforming an image into a data set, that is into descriptors of the image or of the 'objects' which have been depicted in the image. Such features can be used to quantify the image content or to perform a kind of pattern recognition. It should be emphasized that, at this level, the last three steps (preprocessing, processing, and analysis) are not always clearly separated. Preprocessing was previously often qualified as 'cosmetic' processing, that is a kind of processing where the aim is mainly to improve the visual appearance of images. However, this is not the unique purpose of these operations. Similarly, image processing and analysis are often mixed. This is the case with mathematical morphology, for instance, where the same operations can be used for both purposes.

Finally, after image analysis methods have been devised, a statistical approach has also to be defined for estimating the significance of the results. For instance, the number of images which have to be processed before any conclusion can be drawn must be estimated carefully.

The scheme described above concerns the processing/analysis of individual images. It must be stressed that modern microscopes can also provide image series in addition to single images. Different types of image sequence are now produced

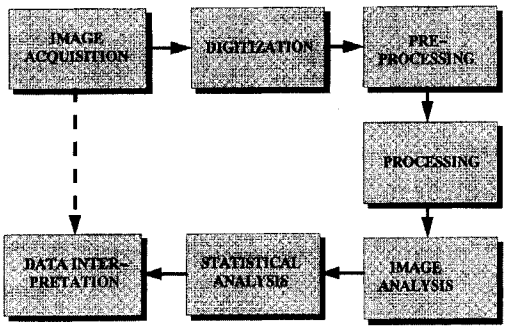


Figure 1. Although we would like to go directly from the image acquisition to data interpretation, it is generally necessary to follow a much longer path, which includes different activities described by the term image processing.

more or less routinely: time series (video-microscopy), spectral series (electron energy loss, X-ray and Auger imaging), depth series (serial sections, depth profiling in SIMS), tilt series (electron or X-ray microscopies), focus series, etc. The combination of one or several of these parameters allows the recording of data in a large dimensional space. Therefore, these complex data sets cannot, in general, be interpreted and handled without the help of a computer. Specific methods have now been developed for the processing and analysis of such image sequences. These methods include, among others, the three-dimensional reconstruction (from serial sections, stereoscopy, or tilt series), the parametric imaging method (modelization approach), and the multivariate statistical analysis.

2.2 Image Preprocessing

Image preprocessing can be considered as an attempt to overcome some of the limitations of the imaging instrument, which have not been accounted for at the acquisition level. (Of course, it must be stressed that everything that can be done at the specimen preparation level or at the experimental imaging level must be done – the less the computer is solicitated, the better.)

Two main concepts are involved at the preprocessing stage: contrast and signal-to-noise ratio (SNR). The contrast of an image is a rather intuitive concept and need not be defined. The SNR is the ratio of the deterministic (or repetitive) contribution to an image over the stochastic (or nonrepetitive) contribution (noise). The ratio of the signal variance to the noise variance, for instance, can be used as a definition of the (quadratic) SNR (QSNR). Noise originates either from electronic fluctuations (detector) or from the quantum uncertainties associated with low-count signals. Although the SNR can often be improved at the experimental level by increasing the acquisition time, it is sometimes necessary to improve it a posteriori, in order to avoid prohibitive acquisition times.

Image preprocessing methods were developed at the early stages of digital image processing. At the beginning, these methods were what we now call *global methods*. This means that they perform the same function at any place in the image, whatever the local image content. Although this approach can be useful, it is often insufficient, as it does not modify fundamentally the image content. For this reason, these methods were criticized for doing mainly a cosmetic job (i.e. they render the images more pleasant to look at, but they do not help very much in the extraction of useful information). Nowadays, new methods (called *local*

methods) have been developed which do a different job according to the local content (contrast and SNR) of images.

2.2.1 Global Methods for Image Preprocessing

An image can be thought of as a collection of pixels along a square (or rectangular, or hexagonal) arrangement (x, y), or a collection of gray levels (GLs) or colors. In the former space, we speak of an 'image space'. It can be described either as a real space (coordinates x, y) or as a reciprocal space (the frequency space: coordinates x^{-1}, y^{-1}). In the later case, we speak of a 'parameter space'. Global image processing can be performed either in the image space or in the parameter space.

2.2.1.1 Examples of Global Image Preprocessing in Image Space or Image Frequency Space

Most of the processing methods in this group are related to the concept of *filtering*. We must be aware of the fact that any image is in fact the superimposition of several components: the low frequencies, which correspond to the slowly varying components of the image (general shape of objects and regions); the intermediate frequencies, which correspond to details of the objects or regions; and the high frequencies, which correspond to rapidly varying parts of the images (i.e., to edges, very small details and also to noise), which in general vary from pixel to pixel. These different components can be visualized in the image spectrum, which can be obtained either by diffractometry on an optical bench or by computing the power spectrum density (squared modulus of the image Fourier transform). Figure 2 shows two different images (one with a mostly low frequency content and one with a high frequency content) and their respective power spectrum density.

Since the different frequency components of an image can be identified, they can also be selectively modified. This is called frequential filtering. Applying a high-pass filter (which consists of lowering the amount of low frequencies in the image spectrum and then returning to the image space by an inverse Fourier transformation) results in contrast improvement, as the weight of details and edges is increased. However, this is at the expense of SNR, because the high-frequency noise components are also enhanced. Applying a low-pass filter (which consists of lowering the relative amount of high frequencies in the image spectrum before computing the inverse Fourier transform) results in an improvement of SNR, as the importance of high-frequency noise components is decreased. This approach, however, is at the expense of contrast and apparent resolution, because the high frequencies of edges and details are also weakened. These two approaches are exemplified in Fig. 3.

Applying a band-pass filter (which consists of selecting specific regions of the image frequency space) may result in contrast enhancement and SNR improvement, as selected frequencies can be reinforced at the expense of others. It must be noted, however, that this procedure, which is often used with images of

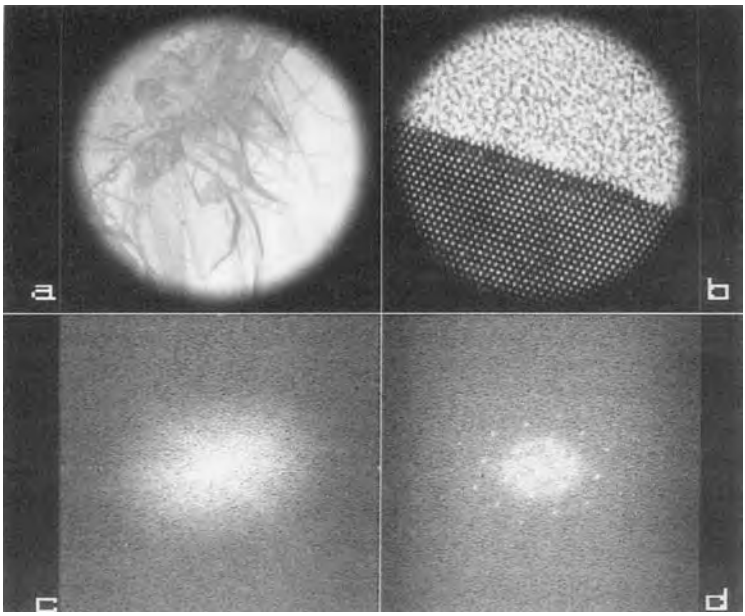


Figure 2. Illustration of the frequency content of an image. (a) Scanning electron microscope image of a whisker; this image contains mainly low frequency components and no periodicities. (b) Transmission electron microscope image of an interface between an amorphous structure and a crystalline structure; this image contains more high frequency components and periodicities. (c) Power spectrum density (squared modulus of the Fourier transform) of (a) displayed on a logarithmic scale. (d) Power spectrum density of (b) displayed on a logarithmic scale; note that it is necessary to apply a circular mask to the image in order to avoid artifacts in the Fourier spectrum.

periodic specimens, may also be dangerous, even if the problems associated to sharp cut-off filters are carefully avoided (in the extreme case, the filtered image of a nonperiodic specimen may appear as periodic, with a period corresponding to the central frequency of the band-pass filter).

Frequential filtering is not the only way to perform these operations, because multiplying by a filter function in the image frequency space is equivalent to performing a convolution in the image space. This process is illustrated in Fig. 4. In practice, the size and the content of the convolution kernel have to be defined. We then have to scan the image to be processed with a moving window of the same size as the kernel. The result (at the center of the moving window) is thus obtained by weighting the image content by the coefficients of the kernel. Each kind of filtering (low-pass, band-pass, high-pass filtering, first order and second order derivatives, Laplacean, etc.) is characterized by a specific convolution kernel.

Nonlinear filters can also be used. For instance, the median filter is often used to improve SNR: it consists of replacing the gray level at the center of the moving window by the median of the values in the window.

2.2.1.2 Examples of Global Image Preprocessing in Parameter Space

When an image is characterized by only the gray level (or color) values and not by the pixel coordinates, at best we have to compute the gray level histogram,

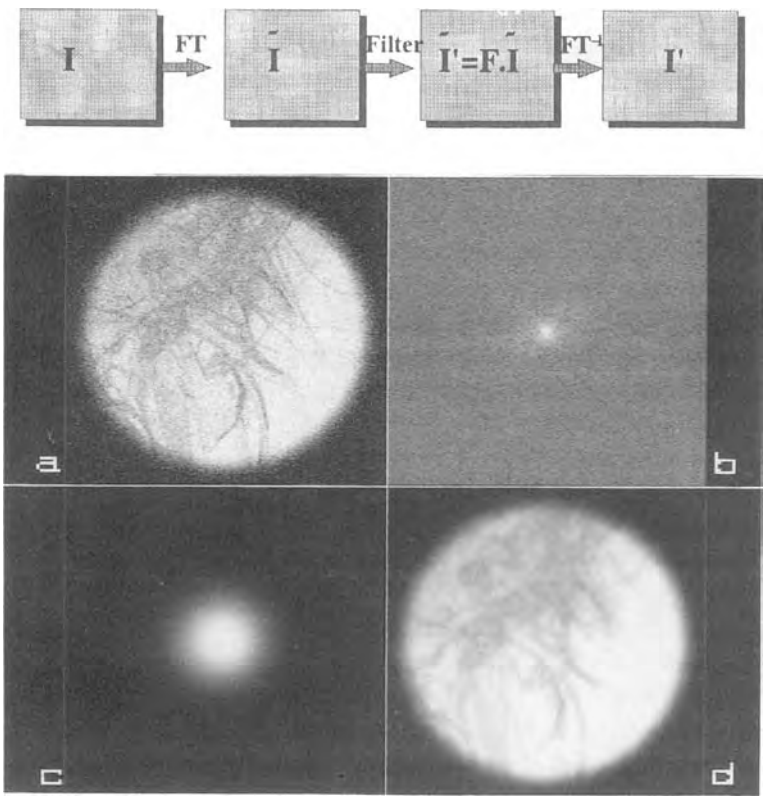


Figure 3. Schematic diagram of the procedure for filtering in the frequency space. (a) Original noisy image. (b) Power spectrum density of (a); logarithmic scale. (c) Low-pass filter applied to the image spectrum. (d) Filtered image; the noise is largely suppressed, but at the same time blurring occurs.

which represents the first-order statistics. (Note that higher order statistics can also be computed, see Sec. 2.4.3 in this Chapter.)

When the gray level histogram is built, it is possible to manipulate it. Replacing the original gray levels by the new ones results also in a new (processed) image. The different possibilities include: video inversion, histogram equalization, histogram hyperbolization, low gray level enhancement (logarithmic transfer function), high gray level enhancement (exponential transfer function), and false color rendering.

This classical approach for gray level manipulation, which is detailed in many textbooks for image processing [1–4], is illustrated in Fig. 5.

2.2.2 Local Methods for Image Preprocessing

The global methods for SNR improvement or contrast enhancement described in the previous paragraph can be found in any commercial software package for image processing. However, these packages suffer from several drawbacks. During the past 10 years, new algorithms have been developed in order to overcome the drawbacks but, unfortunately, they

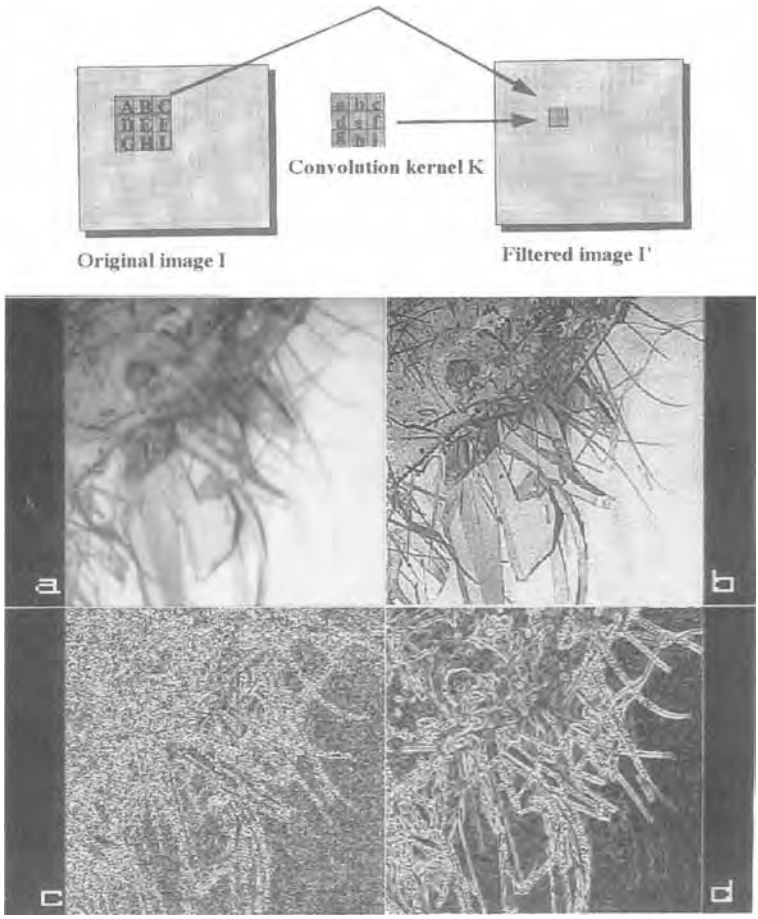


Figure 4. Illustration of the filtering principle according to a convolution process. The result obtained for a position of the moving window is obtained by weighting the local content of the original image I (A, B, C, etc.) by the content of the convolution kernel k (a, b, c, etc.): $I' = I * k = A \cdot a + B \cdot b + C \cdot c + \cdots + I \cdot i$. (a) Smoothing with a 7×7 Gaussian kernel (the original image is that shown in Fig. 5b). (b) Contrast enhancement with a 5×5 high-pass kernel. (c) Edge detection (Kirsch kernel) applied to the original image; the result of this high-pass filtering is much too noisy to be useful. (d) Same edge detection as applied to the smoothed image in (a).

are not included in commercial software. We describe here two examples of such algorithms.

2.2.2.1 Example of Algorithm for Local Contrast Enhancement

The purpose of contrast enhancement is often to increase the detectability of objects against a background (see Sec. 2.3 in this Chapter). Global contrast enhancement does not help very much in this purpose. This can only really be done by local contrast enhancement, as was anticipated by Gabor [5] as early as 1965. An example of such a method was suggested by Beghdadi and Le Negrate [6]. The local contrast is defined for each pixel as

$$C(x,y) = \frac{|I(x,y) - E(x,y)|}{I(x,y) + E(x,y)} \quad 0 \leqslant C \leqslant 1 \tag{1}$$

where I is the pixel gray level and E is the mean gray level value of edges in a local window centered on (x,y) (Fig. 6). Once this local contrast is defined, it can be increased replacing the contrast by

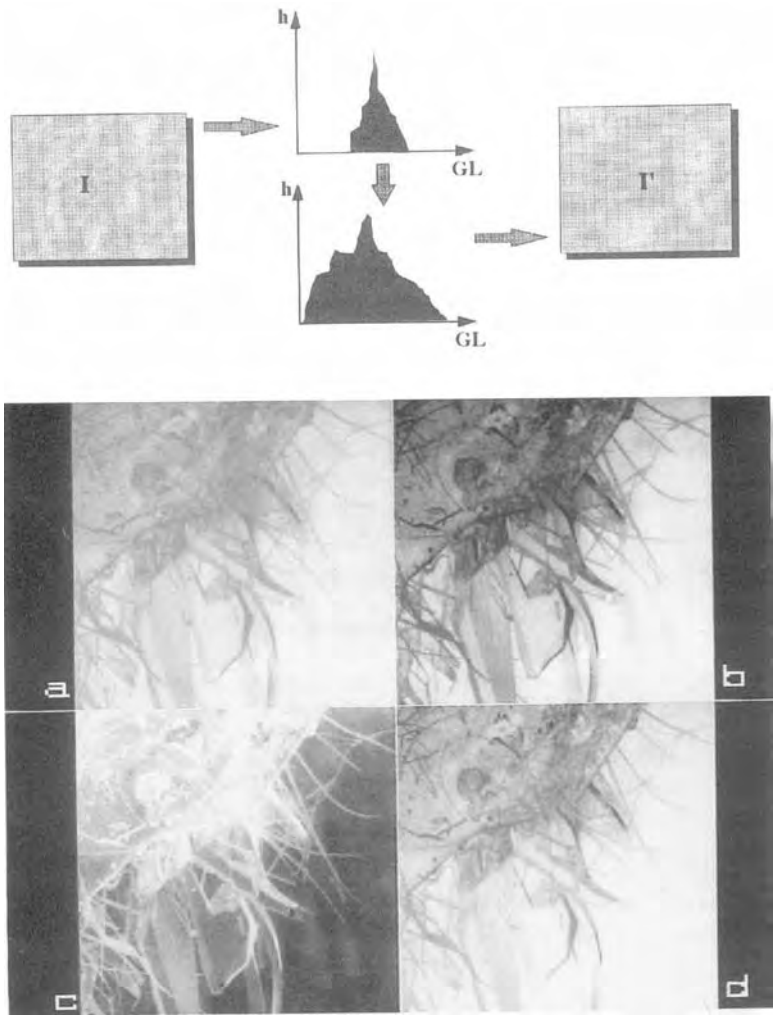


Figure 5. Illustration of global image preprocessing by the histogram-manipulation approach (here, histogram stretching is used as an example). (a) Initial image of whiskers with low contrast. (b) Image resulting from histogram stretching. (c) Contrast inversion of (c). (d) Anamorphosis (logarithmic transfer function) of (a).

$C' = C^n$ $(n < 1)$. From this new contrast, the new gray level value can be computed as

$$I'(x,y) = E(x,y)\frac{1 - C'(x,y)}{1 + C'(x,y)}$$

$$I(x,y) \leqslant E(x,y) \tag{2a}$$

$$I'(x,y) = E(x,y)\frac{1 + C'(x,y)}{1 - C'(x,y)}$$

$$I(x,y) > E(x,y) \tag{2b}$$

This procedure has the attractive property of reinforcing the histogram bimodality; that is it reinforces the appearance of two (or several) distinct peaks (or modes), revealing distinct features.

2.2.2.2 Example of Algorithm for Improving the Signal-to-Noise-Ratio

As described above, the SNR can be improved by smoothing, but this is at the expense of contrast and resolution. SNR

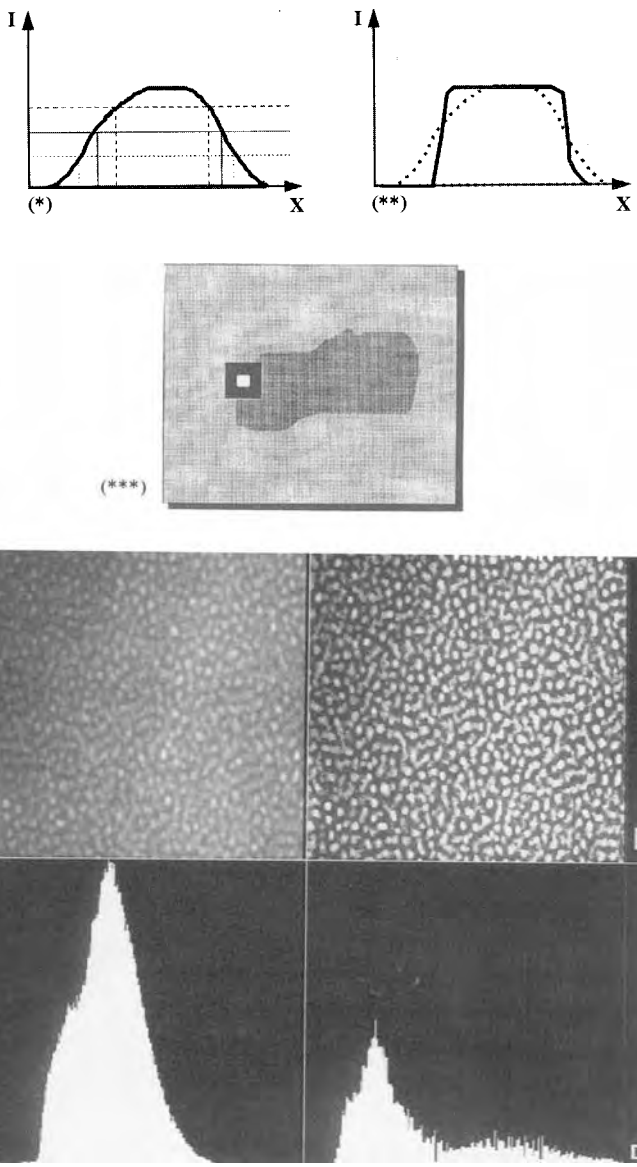


Figure 6. Illustration of a local contrast enhancement procedure. (*) Schematic diagram of a profile across an object: the contrast is weak and thus, for different values of the threshold, different values of the object area (or perimeter) are obtained. (**) The local contrast enhancement procedure helps to define steeper edges from which the binarization procedure is much less sensitive to the threshold value. (***) Illustration of the local window procedure: within this window, the average value of the edge gray levels is computed as:

$$E(x,y) = \frac{\sum_{x',y' \in W} I(x',y') \cdot G(x',y')}{\sum_{x',y' \in W} G(x',y')}$$

where G represents the gray level gradient magnitude. If the pixel at the center of the window has a gray level value lower than E the gray level is decreased; otherwise, it is increased. This results in contrast enhancement. (a) Image of defects observed in silicon after ion implantation. (b) The result obtained after local contrast enhancement (window size 31 pixels). (c) Histogram of (a). (d) Histogram of (b). The local contrast enhancement procedure has the attractive property of reinforcing the histogram bimodality. (Unpublished. Courtesy of J. Fauré, University of Reims, and S. Simov, Institute of Solid State Physics, Sofia.)

improvement without the loss of too much contrast and resolution can be obtained if smoothing is performed on approximately constant gray-level areas while smoothing is avoided on regions containing details and edges. Anisotropic diffusion [7] is one of the procedures which has been developed for this purpose: it is an iterative

procedure which consists of computing the image gradient, performing weighted smoothing (with weights inversely proportional to the gradient magnitude), and refining the gradient computation, etc.

Other approaches to this problem are: Unser's filter [8], Maeda and Murata's filter [9], and kriging [10, 11].

2.3 Processing of Single Images

Processing an image means transforming an image into another one which is easier to interpret or easier to analyze quantitatively. To illustrate these two purposes, we have chosen to describe two different groups of image processing: image restoration and image segmentation.

2.3.1 Image Restoration

Image restoration consists of taking into account the instrument characteristics and the experimental conditions which sometimes do not make the image a faithful representation of the object function, but instead a degraded representation. When these microscope and imaging characteristics are known, or can be determined, a restoration of the image can be attempted. It must be emphasized that, in some cases (e.g., an electron microscope working at the ultimate resolution), the restoration of the object function at the exit plane of the specimen is the main goal of image processing, as it is a prerequisite to any image interpretation.

The different kinds of degradation of the object function during image formation include: geometric distortion, uneven illumination, detector nonlinearities, aberrations, and additive and multiplicative noise. Most of these degradations can be eliminated (at least partly) a posteriori, provided they can be characterized [12]. Below we illustrate some, but not all, possibilities.

2.3.1.1 Restoration of Linear Degradations

A linear system can be described by an impulse response h in the real space or a transfer function H in the frequency space

$$\mathrm{i}(x,y) = \mathrm{o}(x,y) * \mathrm{h}(x,y) + \mathrm{n}(x,y)$$

or

$$\mathrm{I}(x^{-1},y^{-1}) = [\mathrm{O}(x^{-1},y^{-1}) \cdot \mathrm{H}(x^{-1},y^{-1})] + \mathrm{N}(x^{-1},y^{-1}) \qquad (3)$$

where o stands for the object function, i for the image, n for additive noise, O, I and N for their Fourier transforms and $*$ for the convolution product. The impulse response and the transfer function contain a contribution of the instrument characteristics (aberrations, objective diaphragm aperture, etc.) and of the imaging conditions (defocus value for instance).

Restoration by a simple inverse filtering procedure

$$\hat{\mathrm{O}}(x^{-1},y^{-1}) = \mathrm{I}(x^{-1},y^{-1}) \cdot \mathrm{H}^{-1}(x^{-1},y^{-1}) \rightarrow \hat{\mathrm{o}}(x,y) \qquad (4)$$

where $\hat{}$ indicates an estimated function, does not provide a correct solution, because the frequencies at which $\mathrm{H} = 0$ are amplified infinitely, and this results in a degraded SNR.

A Wiener filter gives a much better solution, as the second term in the denominator (a noise-to-signal ratio) prevents the filter from becoming infinite [13]:

$$\hat{\mathrm{O}}(x^{-1}, y^{-1}) = \frac{\mathrm{I}(x^{-1}, y^{-1}) \cdot \mathrm{H}^*(x^{-1}, y^{-1})}{\left[\mathrm{H}^2(x^{-1}, y^{-1}) + \dfrac{\mathrm{N}^2(x^{-1}, y^{-1})}{\mathrm{O}^2(x^{-1}, y^{-1})}\right]} \quad (5)$$

The difficulty arises in this noise-to-signal ratio, which has to be estimated independently.

An alternative solution is given by iterative methods which consist of: (a) estimating the object function $\hat{\mathrm{o}}_k(x, y)$, the first estimate $\hat{\mathrm{o}}_0$ being the image itself; (b) computing the image which such an object would give in the experimental conditions described by $\mathrm{h}(x, y)$; and (c) comparing the computed and the experimental image and updating the object function $\hat{\mathrm{o}}_{k+1}(x, y)$ according to the difference and to some additional constraints. Note that this procedure converges for strictly positive transfer functions, while in the other cases a reblurring procedure must be devised [14].

2.3.1.2 Restoration of Partly Linear Degradations: Very High Resolution Electron Microscopy

In the previous section we assumed that image formation can be characterized by a linear transfer function. Unfortunately, this is not always the case. In transmission electron microscopy, for instance, the image intensity (and not the image amplitude) is recorded (for books specifically devoted to image processing in electron microscopy, see [15, 16]).

This intensity can be written as

$$\mathrm{i} = (1 + \mathrm{o} * \mathrm{h})^2 = 1 + 2\mathrm{o} * \mathrm{h} + (\mathrm{o} * \mathrm{h})^2 \quad (6)$$

Thus, even for weak phase objects, the last term is nonlinear and prevents any restoration based on the linear system theory. Restoring the object wavefunction o therefore necessitates more sophisticated schemes, including: focus series, tilt series, and holography (see this volume, Chap. IV, Sec. 1.8). Owing to the limited space, we restrict ourself to a brief description of one of these approaches: the focus series approach.

This approach to object wave restoration was suggested as early as 1973 [17], but within the framework of the linear imaging theory. Later, image reconstruction from a focus series was considered as a nonlinear process and tackled by a recursive least-squares formalism. Recent developments [18, 19] can be described as: (a) firstly, preliminary reconstruction with a fully linear procedure; (b) secondly, a few iterations where the nonlinear terms are treated as a perturbation; and (c) thirdly, a fully nonlinear procedure, according to the paraboloid method (PAM). The updated electron wave is obtained from the previous estimation by

$$\hat{\mathrm{o}}^{(k+1)} = \hat{\mathrm{o}}^{(k)} + \gamma^{(k)} s^{(k)}[d(\hat{\mathrm{o}}^{(k)})] \quad (7)$$

where $d(\hat{\mathrm{o}}^{(k)})$ is the difference wave computed from least-squares minimization, $s^{(k)}$ is the best search direction determined by a conjugate gradient approach, and $\gamma^{(k)}$ is a feedback parameter (gain).

Applications of this kind of approach concerns studies at the atomic resolution level. The amplitude and phase of the object function are reconstructed separately and can be used to determine the

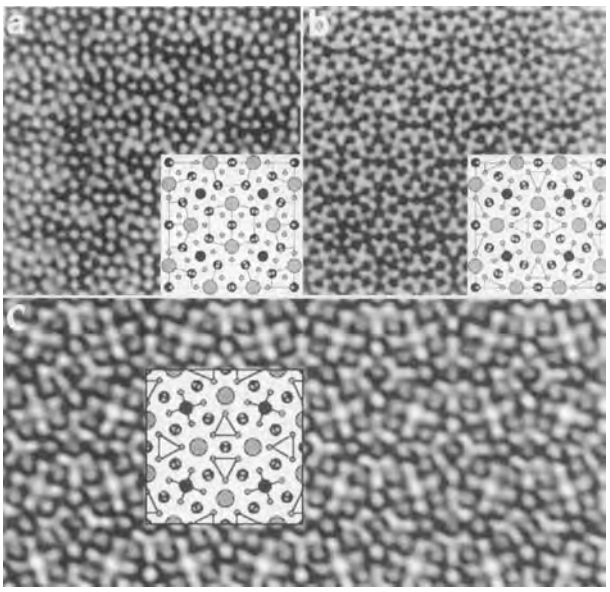


Figure 7. Example of restoration of the object wavefunction in high resolution electron microscopy (Courtesy of D. Van Dyck and M. Op de Beeck, RUCA, Antwerp): (a) Amplitude of the reconstructed exit wavefunction starting from 20 images of $Ba_2NaNb_5O_{15}$. Mainly the heavy atoms are visualized. (Thickness 12 mm, microscope CM20FEG–SuperTwin). (b) Phase of the same reconstructed exit wavefunction mainly showing the light atoms. (c) Structure reconstruction towards the 1S channelling eigenstates of the different columns, showing all atomic columns in one image. (Structure models inserted).

'true' atomic positions. An example of reconstruction is given in Fig. 7. It concerns the structure of $Ba_2NaNb_5O_{15}$ where the light atoms are revealed in the phase and the heavy atoms in the amplitude.

Numerous variants of this approach (tilt series reconstruction [20], stochastic algorithms [21], maximum entropy reconstruction [22], and a crystallographic approach [23, 24]) have been suggested, and have provided encouraging results. However, they require further study before they can be used routinely.

2.3.1.3 Example of a Completely Nonlinear Restoration: Near-Field Microscopies

In scanning tunneling microscopy and atomic force microscopy, image formation does not follow the scheme of imaging instruments with lenses (see this Volume, Chap. VII, Sec. 1). For corrugated surfaces, the image contrast is governed by both the shape of the specimen surface o and that of the tip s. It can be shown [25] that the relationship is of the form:

$$i(x, y) = \max_R [o(x', y') - s(x' - x, y' - y)] \tag{8}$$

This relation is clearly nonlinear and it has been shown [26] that the image function can be considered as the dilatation of the specimen surface function by a three-dimensional structuring element, the tip (see Sec. 2.3.2.5 in this Chapter for a definition of the dilatation operation and of the structuring element).

As image formation is not a linear operation, we cannot expect a linear operation to help in restoring the object function (the local specimen height). In fact, we can show that the best restoration can be obtained through the image erosion by the tip shape as a three-dimensional structuring element

$$\hat{o}(x, y) = \min_R [i(x', y') + s(x - x', y - y')] \tag{9}$$

The tip shape itself can be estimated when an image of a well-known object is made:

$$s(x,y) = -\min_{R} [i(x',y') - o(x'-x, y'-y)] \quad (10)$$

Owing to the specific properties of the dilatation/erosion operations, we can also show that some estimate of the tip parameters can be obtained on the basis of the image alone, that is without any reference object (when the tip shape is known) [26].

2.3.2 Image Segmentation

Image segmentation is that part of image processing which consists of partitioning the whole image into different constituents. Depending on the type of image, this can mean one (or several) object(s) over a background or different regions. While this task is easy to perform for the human visual system (because it is associated to the observer's brain), it is a very difficult task for a computer, which does not 'know' the purpose of the segmentation (Fig. 8).

Many attributes of the image can be used in trying to perform an automatic segmentation, including: the gray levels; the gray-level gradients; the gray levels and the concept of connectivity; and the gray levels, gradients and connectivity together. This explains why hundreds of methods have been tried for performing this task (for a review see [27]).

2.3.2.1 Segmentation on the Basis of Gray Levels Only

As stated above, when only gray levels (or colors) of an image are considered, we

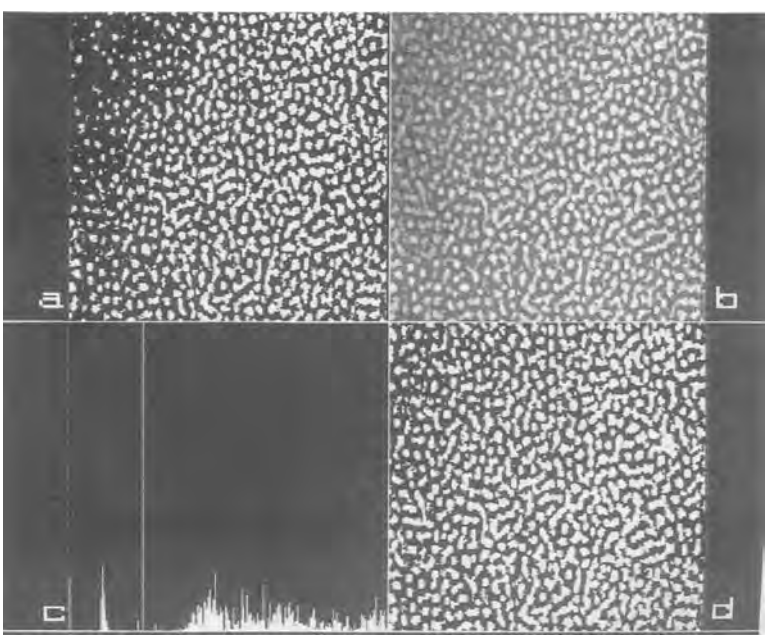


Figure 8. (a) Illustration of the difficulty in segmentation using the histogram alone. The image displayed in Fig. 6b was binarized according to the histogram valley position. The segmentation is far from perfect, despite the histogram bimodality. (b) Result of applying the region-growing algorithm to the image in Fig. 6b (the threshold for growing was set at 30 gray levels). (c) Histogram corresponding to (b). Three main peaks on the left correspond to the background, while the others correspond to different granular structures. (d) The result obtained when binarizing (b) is much better.

have to work with the histogram. A necessary condition to perform segmentation on the basis of the histogram alone is that it is bimodal (or multimodal). If this is so, it can be assumed that each group of objects (or each group of regions) is characterized by a peak in the histogram. Thus, it is necessary to define the thresholds between these peaks and to attribute the pixels to the different regions according to their gray-level values (compared to the segmentation thresholds). A number of methods have been suggested in order to find, automatically, the thresholds of multimodal histograms (for reviews see [28, 29]). Since the different methods rely on different assumptions concerning the statistical characteristics of an image, none of them are expected to give the best solution in any situation. Furthermore, the histogram multimodality, though a necessary condition, is not expected to be a sufficient condition: it is easy to build examples where the peaks in a histogram are not representative of distinct regions. Thus, the segmentation on the basis of the histogram is often rather questionable.

2.3.2.2 Segmentation on the Basis of Gray-Level Gradients

When we try to go beyond the simple concept of gray levels, it is essential to define a strategy for segmentation. Two strategies have previously been defined. The first consists of defining objects (or regions) as sets of connected pixels which possess some kind of attribute in common. The attribute can be a gray level, a color, or a texture. This concept of attribute homogeneity leads to the algorithm of 'region growing' (see below). The other strategy consists of defining contours of objects (or boundaries of regions) as pixels for which there is a strong discontinuity of an attribute. These two approaches are in fact complementary since, when regions are defined, their boundaries are also known, and when closed boundaries are defined regions inside these boundaries are also known. But the intermediate steps differ greatly for these two strategies.

The discontinuity-based strategy consists first of computing either the gray-level gradient or the gray-level Laplacean (an edge can be identified either by the local maximum of the gradient amplitude or by the zero-crossings of the second-order derivative).

Despite the fact that derivative operations are ill-conditioned (i.e., are very sensitive to a small amount of noise), very efficient digital filters are now available for performing such computations [30, 31]. However, this first step is not sufficient, because 'edges' obtained in this manner are noisy, unclosed, etc. Therefore, a number of postprocessing operations, such as edge closing, edge following, and hysteresis [32], are needed in order to obtain the real boundaries of regions.

2.3.2.3 Segmentation on the Basis of Gray-Level Homogeneity and the Concept of Connectivity

As described above, an object or a region can be defined as a connected set of pixels for which some attribute is homogeneous. From this definition, an algorithm can be defined: from the seed of a region, consider connected pixels and incorporate them in the same region if the difference between

the pixel gray value and the average gray value of the region is smaller than a predefined threshold. Otherwise, the pixel serves as a seed for a new region. This region-growing algorithm is easy to understand and easy to implement. However, it also suffers from some drawbacks: (a) the threshold has to be defined, which is not an easy task, so several thresholding values have to be tried in order to choose the best segmentation, according to some criterion of what is a 'good' segmentation; and (b) the result is very sensitive to the choice of the seeds (i.e., to the order in which the pixels are processed).

This region-growing method is sometimes used in conjunction with a region-splitting method (split-and-merge algorithm [33, 34]).

2.3.2.4 Segmentation on the Basis of Gray Levels, Gradients, and Connectivity: Functional Minimization

Owing to the difficulties associated with the methods just described, and also to the lack of criterion for good segmentation, a more recent approach consists of associating a cost function with any segmented image. The idea is to associate two terms: the first guarantees that the different regions remain sufficiently close to the original image (fidelity principle) and the other states that the boundaries of the regions must be as short as possible (simplicity principle). The cost function (also called the energy function) can then be written as:

$$E = \sum_{R} \sum_{x,y \in R} |\mathrm{i}(x,y) - E_R[\mathrm{i}(x,y)]|^2 + \alpha L \tag{11}$$

where E_R is the average gray level of region R, L is the total length of boundaries and α is a weighting coefficient proportional to the size of the smallest desired region [35].

The strategy of segmentation consists of minimizing this energy function. From this general strategy, several algorithms can be defined: either approximate contours are drawn interactively and then (automatically) deformed iteratively until a minimum energy is attained ('snakes' approach), or small homogeneous regions are merged until a minimum of the energy function is also obtained (conditional region-growing approach).

2.3.2.5 Mathematical Morphology

In addition to all the approaches described above, an important class of procedures deals with the concept of neighbor relationship. These procedures form the basis of mathematical morphology [36–38], which is concerned with operations that respond to a feature shape (segment, disc, etc.) called a structuring element. The applications of mathematical morphology are not limited to image segmentation, but concern also image preprocessing (cleaning) and image analysis [39]. The basic tools are:

(a) *Erosion.* At every pixel position, the result is the minimum gray-level value over the part of the image covered by the structuring element, placed at this position.
(b) *Dilatation.* The maximum gray level value is selected instead of the minimum value.

From these two basic operations, many others can be defined, such as:

(c) *Opening.* Erosion followed by dilatation.
(d) *Closing.* Dilatation followed by erosion.
(e) *Morphological gradient.* Eroded image subtracted from the dilated image.
(f) *Top-hat transformation.* Open image subtracted from the original image; this allows the detection of small bright particles (or details, or spots) over an uneven background.

What could be the most powerful tool for image segmentation is the watershed transformation [40, 41], which can be considered as a region-growing approach starting from markers of the regions: the regional minima. From these minima, a flooding process is realized, which results in different catchment basins separated by the watershed lines.

2.4 Analysis of Single Images

Once an image is split (automatically or interactively) into several objects or several regions, image analysis can take place. This consists of computing features of the different constituents.

2.4.1 Object Features

A nonexhaustive list of features includes:

(a) *Geometrical features.* Area A, perimeter P, Feret diameters (sizes of the smallest rectangle containing the object), Euler number (number of objects minus number of holes), intercept numbers, etc.
(b) *Shape features.* Shape can be characterized by a number of methods: (i) global shape factors such as form factor $(P^2/4\pi A)$, eccentricity (or ellipticity), which characterize the elongation of the object; (ii) geometrical moments [42]; and (iii) coefficients of the Fourier series (development of $\boldsymbol{p}(s)$, where $\boldsymbol{p}$ is the vector radius computed from the center of mass, and s is the normalized arc length) [43].
(c) *Densitometric features.* Provided a proper calibration can be made, the integrated gray levels within object can be related to one of the physical characteristics of that object (density, absorption coefficient, molecular mass, etc.).
(d) *Texture and fractal parameters.* (See Sec. 2.4.3 of this Chapter).
(e) *Spatial distribution features.* See a discussion in the literature [44].

2.4.2 Pattern Recognition

Pattern recognition is also a part of image analysis. Several aspects can be encountered.

Pattern recognition may consist of detecting objects of known shape within images, without performing image segmentation. One of the most used approaches in this context is the generalized Hough transform. The Hough transform was originally designed for detecting straight lines. It consists of mapping each pixel of the original image into a parameter space and characterizing the object to be detected; for example, for a circle the parameter space is composed of the center coordinates and the radius. When such an

object effectively exists in the image space, an accumulation point appears in the parameter space. In electron microscopy, the Hough transform was used to detect automatically lines in Kikuchi patterns [45]. The Hough technique has been extended to the detection of circles, ellipsoids, polygons or objects of any known shape [46].

Pattern recognition may consist of the detection of any object, characterized by its gray-level pattern. In this case, a *matched-filter* technique can be used, which is generally implemented as a cross-correlation technique [47]. The object to be detected is placed at the center of an empty image of the same size as the real image where it is to be searched for. The two images are then cross-correlated:

$$C(\Delta x, \Delta y) = \frac{E[i_1(x,y) \cdot i_2(x+\Delta x, y+\Delta y)] - E[i_1(x,y)] \cdot E[i_2(x+\Delta x, y+\Delta y)]}{\sigma(i_1) \cdot \sigma(i_2)} \quad (12)$$

where E is the expectation value and σ the standard deviation. It can be shown that peaks of the $C(\Delta x, \Delta y)$ function appear in places where the objects are located.

Pattern recognition may consist of recognizing or classifying objects on the basis of the features which have been computed from them. Recognizing objects assumes that a bank of objects (with their associated features) already exists. The problem is then one of *supervised* pattern recognition, for which numerous methods exist, ranging from the computation of the smallest distance to neural network approaches. Classifying a set of objects without reference to previously known classes is a problem of *unsupervised* pattern recognition (or automatic clustering) [48].

2.4.3 Image Analysis without Image Segmentation

The usual procedure of image processing (segmentation) followed by image analysis (computation of object features) is sometimes inoperative because the whole image is either composed of only one region, which has to be characterized globally, or is composed of a very large number of very small or overlapping objects that cannot be isolated and characterized individually. In these two extreme cases specific tools have to be used for performing image analysis. Some examples of these tools are described below.

2.4.3.1 Texture Analysis

The problem of texture recognition is one of the problems for which the human visual system is much more efficient than any computer system. However, many tools are now available. Texture may be recognized through different attributes, including graininess, directionality, uniformity, and elongation. These features can be characterized and quantified with concepts generally related to second-order statistics, that is statistical parameters relating the gray level of one pixel to the gray level of another pixel (a distance d apart). Such parameters include co-occurrence matrices and their descriptors [49], output amplitudes of real space filters [50, 51], Gabor frequency space filters [52] and wavelet decomposition.

Pixels (or groups of pixels) of textured images can thus be described by vectors of attributes and thus subjected to pattern recognition or pattern classification (texture segmentation).

2.4.3.2 Fractal Analysis

The pioneering work of Mandelbrot [53] has shown that natural scenes often exhibit self-similarity, that is they have similar behavior on different scales. This property can be used to characterize images which otherwise appear very chaotic. Here, again, many different methods have been suggested for estimating fractal attributes, the most used of which is the fractal dimension D:

(a) *Area–perimeter method.* Following the suggestion of Mandelbrot, the fractal dimension can be estimated by performing image binarization at different thresholds (pixel values below the gray level threshold are set to zero, while pixel values above the threshold are set to one). For each threshold, the total area A and total perimeter P of the thresholded objects are computed. The linear regression of the $\log(P)$ versus $\log(A)$ data provides an estimation of D. An application to STM images is given in [54].

(b) *Power spectrum method* [55]. For a purely fractal object, the power-spectrum density is proportional to the spatial frequency at the power of a parameter H connected to D: $P(|f|) = |f|^{-2(H+1)}$. Thus, a $\log(P)$ versus $\log(|f|)$ regression allows us to estimate $D = 2 - H$.

(c) *Hurst coefficient method* [56]. For each pixel in an image, the first, second, third, etc., neighbors are considered and the variance of the gray levels for each neighborhood plotted as a function of the distance (on a log–log scale). Again, the slope of the line provides an estimation of the local fractal dimension D.

(d) *Blanket method* [57]. The gray levels are considered as the third coordinate of a $z(x, y)$ plot. The surface thus created is iteratively expanded towards the higher and lower gray-level values by computing the maximum and minimum values over a given neighborhood. The surface variation is then plotted (on a log–log scale) as a function of neighborhood size. Again, the slope of the curve gives an estimate of the fractal dimension D. The advantage of this method (which is closely related to the mathematical-morphology approach) is that both the upper and lower sides of the surface can be characterized individually.

Comparative studies are presently being performed in order to evaluate the advantages and inconveniences of the different approaches in specific situations.

2.4.3.3 Stereology

Another way to obtain estimates of certain geometrical quantities without segmenting an image into its different constituents is provided by the stereological approach. More specifically, stereology allows us to estimate quantities in an N-dimensional space from measurements (or countings) performed in a space of dimension $(N - 1)$ [58, 59]. For instance, the ratio between the volume of two compartments (V_A/V_B) can be estimated from their average area ratio (S_A/S_B), measured in random two-dimensional sections through the three-dimensional volume. The area ratio itself can be

estimated from the average ratio of the number of counts (N_A/N_B) of points placed in a systematic (or random) way on to the images of the sections (N_A is the number of points which fall inside structure A, and N_B the number of points within the image of structure B). Several other, but not all, figures of interest can be estimated using similar stereological methods. Increasingly efficient stereological procedures have recently been developed for obtaining less biased estimations, even for anisotropic structures and objects [60, 61].

2.4.4 Mathematical-Morphology Approach to Image Analysis

As for image processing, mathematical morphology provides a large set of specific tools for image analysis. Some details of a limited number of these tools are given below.

2.4.4.1 Granulometry

The case of images composed of a large number of small (possibly overlapping) objects has been mentioned previously. Suppose that we want to compute the size-distribution histogram of these objects. Apart from the possibility of characterizing each object individually, a more fruitful approach consists of working with the whole image. This can be done by opening granulometry: when performing opening with a structural element of size λ, objects with a smaller size than λ are suppressed. Therefore, performing opening with structuring elements of increasing size acts like a sieve. It can be shown [37, 63] that the size-distribution histogram can be obtained as

$$g(\lambda) = \frac{dG(\lambda)}{d\lambda} \tag{13}$$

where $G(\lambda) = [i - O_\lambda(i)]/i$ and $O_\lambda(i)$ is the open image with a structuring element λ (see Fig. 9). It should be noted that opening granulometry can be performed on binary images as well as on gray-level images [62, 63].

This approach of characterizing complex images by a statistical parameter can be largely extended to other parameters. We then reach the field of random image models [64, 65]: from the random sets theory [66], numerous random function (RF) models (such as the Boolean RF model, the dead-leaves RF, the alternate sequential RF, the reaction–diffusion RF or the dilution RF) can be built. The main purposes of these scalar or multivariate models are: to sum up the microstructure in a limited number of parameters; to predict morphological properties; and to provide a means of simulation. They are appliced, for example, in studies of powders, fibers, aggregates, microsegregations, multiphase microstructures and multispectral mapping.

2.4.4.2 Distance Function

Image analysis may concern the spatial repartition of objects within regions. One useful tool in this context is the distance function. Assuming that a region is defined, the aim of the distance function is to define the distance of any interior pixel to the closest boundary. This can be done by eroding the binary region

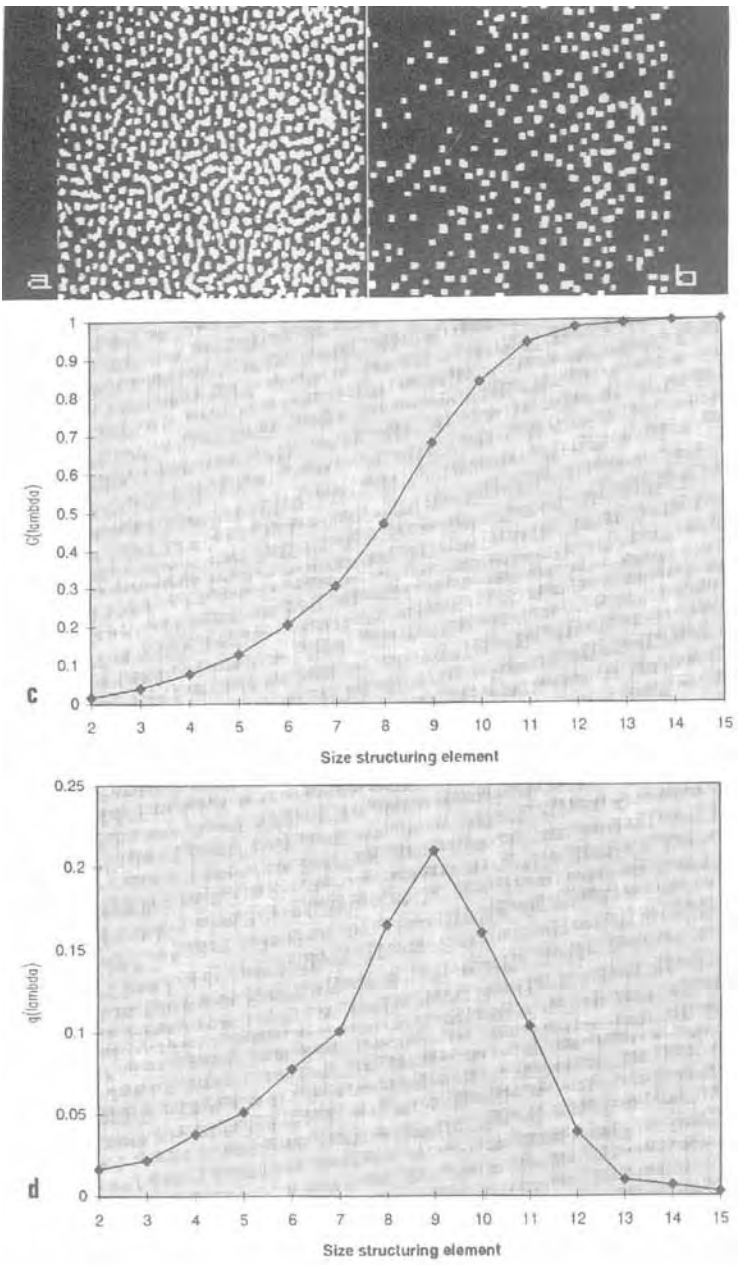


Figure 9. Illustration of the granulometry procedure. Examples of intermediate results obtained after opening the image in Fig. 8d by a structuring element of size (a) 5 and (b) 9 pixels, respectively (images contain 512 × 512 pixels). (c) Curve representing $G(\lambda)$. Curve representing $g(\lambda)$.

iteratively. The distance function is built by assigning to each pixel the number of erosions needed to attain it. Once the distance function has been computed, concentric zones can be defined easily. Computing the watershed of the distance function may also help in separating touching particles.

2.4.4.3 Skeleton

The skeleton of an object (also called the medial axis transform) is the set of pixels which correspond to the midline of that object. It is generally composed of a trunk and branches. It can be described by several attributes (number of branches, end-points, triple points, loops, etc.). It can be shown that the skeleton concentrates the information on the object into this limited set of features, and can thus be used for pattern recognition. It can be computed according to several algorithms (iterative conditional erosion, regional maxima of the distance function).

The skeleton by influence zones (SKIZ) is the skeleton of the surroundings of one or several objects, which can be obtained by negating the binary image. It can be used for separating touching features, or for defining the zones of influence of objects. It is also called the Voronoï diagram (see Fig. 10). A recent extension of this approach is the calypter [67].

2.5 Processing/Analysis of Image Series

As stated in the Introduction, image sequences are increasingly becoming recorded in addition to individual images. We will consider two groups of applications in this very important extension to imaging: the three-dimensional reconstruction of objects and the analysis of a spectral/temporal/spatial image series.

2.5.1 Three-Dimensional Reconstruction

For a long time imaging methods have suffered from the fact that a two-dimensional

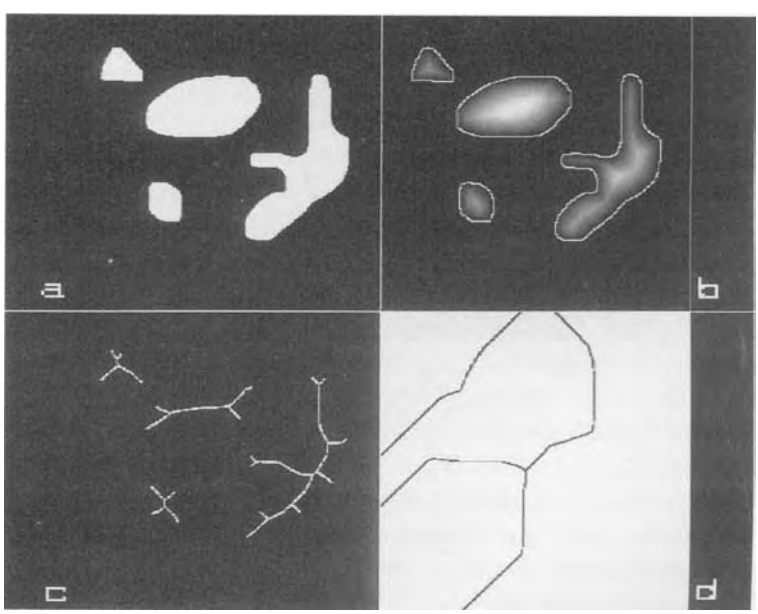


Figure 10. Illustration of some tools of binary mathematical morphology. (a) Binary image containing several objects. (b) Distance function, with the object contours overlayed. (c) Skeletons of the objects. (d) SKIZ of the objects, representing their zones of influence.

image can only give a limited representation of the three-dimensional world. Nowadays, things are changing rapidly, since for most transmission imaging methods tools have been developed for performing the three-dimensional reconstruction. Of course, several images must be recorded for this purpose. Three main groups of methods are available: serial sectioning, stereoscopy, and microtomography.

2.5.1.1 Serial Sections

With this approach, many images of longitudinal sections of the specimen are recorded. The way in which these images are obtained depends on the depth of field of the imaging instrument. In optical or confocal microscopy, it is possible to focus at different depths in the specimen and to record 'optical' sections. In transmission electron microscopy, this is not possible and the only way to obtain serial sections is to cut the specimen ('mechanical' sections). In SIMS, the principle consists of eroding the specimen and recording images of the successive eroded slices.

The principle of three-dimensional reconstruction from serial sections is very simple: it consists of stacking the different digital images into a digital volume, taking into account the height increment (Δz) between images and verifying that no lateral shift occurred during image acquisition.

2.5.1.2 Stereoscopy

Instead of looking 'through' the object, another possibility of recording information relevant to the third dimension is to look at the object from different viewpoints'. This is done in stereoscopy and microtomography. In stereoscopy (mainly used in transmission electron microscopy), two images of the object are recorded for two directions of illumination relative to the specimen [68]. Within these two images (which constitute a stereopair), the distance between two details situated at different depths is different (the difference is called the *parallax*). This coded three-dimensional information must be decoded a posteriori [69]. This can be done either analogically or by digital means. Analog modes consist of making use of the human visual system: one image of the stereopair is shown to the left eye and the other one to the right eye. The human brain reconstitutes the third dimension and the object can be seen in relief. This approach can be carried out using either a stereoscope or computer procedures and a display screen (anaglyph procedure with red/green spectacles, polarized light procedure, etc.).

Digital three-dimensional reconstitution consists of computing the height (z) of any object detail from its coordinates ($x_1, y; x_2, y$) in the stereopair (y is the tilt axis and must be common to both images: the epipolar constraint). When images are simple (i.e., contain only a few well-separated object details), correlation procedures allow us to put these homologous details into correspondence and thus to compute the parallax automatically. For more complicated images, most of the reconstructions that have been made to date in microscopy have been performed interactively: the user has to define (with a graphic cursor) homologous details in the two images, a procedure which is rather tedious. However, according to recent progress described in the image processing

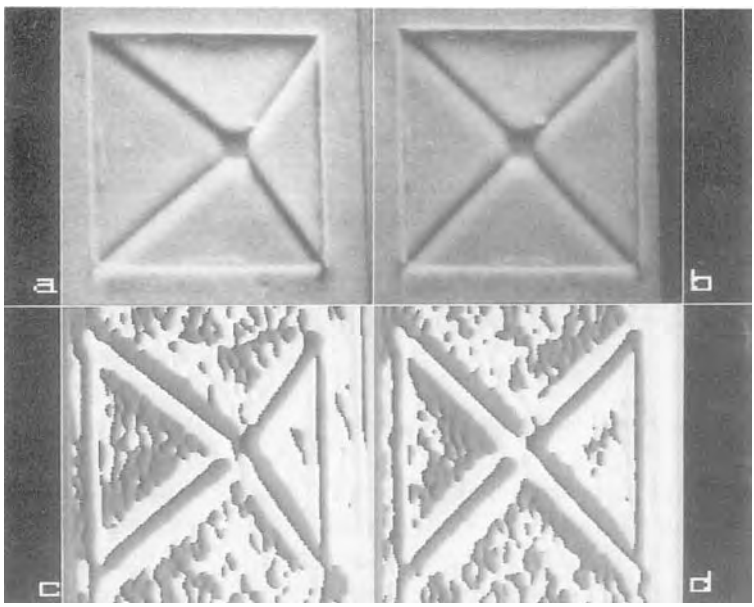


Figure 11. Illustration of three-dimensional reconstruction by stereoscopy. (a, b) Stereopair of an etchpit, recorded in MEB (courtesy of S. Simov, Institute of Solid State Physics, Sofia; for more details see [69]). From the coordinates of specific features (corners), specified manually or detected automatically in the two images, the third coordinates (height) can be estimated, as can the angles between segments and planes. From these angles, the crystallographic indices of the faces can be deduced. (c, d) Illustration of the phase (in the horizontal direction) of images (a) and (b). Comparing the phases of the left and right images allows us to deduce the local parrallax, and to build the elevation map (unpublished results).

literature (see for example the phase method [70, 71] for computing disparity), we can hope that in the near future three-dimensional maps will be built automatically from a stereopair (see Fig. 11).

2.5.1.3 Microtomography

In transmission imaging instruments, which can be equipped with a goniometer stage (transmission electron microscopy (TEM), X-ray microtomographs), it is possible to perform at the microscopic level the equivalent of the medical X-ray scanner. Experimentally, a tilt series (i.e., a large number of images corresponding to as many different orientations of the specimen as possible) has to be recorded. In a sense, microtomography can be understood as a generalization of stereoscopy (as N views of the object are recorded, instead of two), but the philosophy is also different, because the gray-level distribution within the object is restored, instead of only the geometrical position of the details.

Numerous algorithms for performing this reconstruction have been suggested [72–74]; here we give just a few indications about two of them. We consider a single tilt axis geometry (images are recorded when tilting the specimen around a horizontal y axis) and a parallel-beam illumination system (for a description of a cone beam geometry see this Volume, Chap. II Sec. 3). In this configuration, vertical planes (x, z) of the object are reconstructed

from homologous lines in the images. The general reconstruction principle is that of back-projection: since images are projections (and integration) of the object function along the illumination direction, the object function must be a combination of the back-projections of the images along their respective illumination directions. The problem is to know how to combine these different back-projections.

One approach (originating from medical imaging) is the filtered back-projection: the projection/back-projection process can be characterized by a (low-pass) linear transfer function. Thus the 'true' reconstruction consists of a filtered back-projection: during the course of back-projection, a high-pass filter performing the inverse of the (low-pass) projection filtering is applied. A variant of this approach can be used for crystalline materials for which very high resolution three-dimensional reconstructions can be obtained (see e.g., [75]).

Another approach consists of iteratively refining the reconstructed object by comparing experimental images with projections computed from the object reconstructed in the previous step (note the analogy with iterative methods for image restoration; see Sec. 2.1.1 in this Chapter). An example of the application of this approach to X-ray microtomography is given in this Volume, Chap. II Sec. 3. Numerous variants of this approach have been (and still are) suggested, the aim being to reduce as much as possible the imperfections of the reconstruction (the most important of which is its anisotropy, which comes from the fact that the goniometer stages generally allow only partial coverage of the solid angle).

2.5.1.4 Three-Dimensional Display

Whatever the method used for the three-dimensional reconstruction, the next problem is to display the three-dimensional digital volume, a problem which is much less trivial than displaying images. Fortunately, to do this we can take advantage of the advances made in the field of image synthesis. Many methods are thus now available for efficient rendering of three-dimensional images: surface rendering, volume rendering, ray-tracing [76], and other new methods. However, whatever the performance of a three-dimensional static rendering approach, the best way to render the whole information contained in three-dimensional digital data sets, is to perform dynamic rendering (i.e., to produce movies). These movies can now be played easily on computer screens.

2.5.2 Processing and Analysis of Spectral, Temporal and Spatial Image Series

Besides image sequences for image reconstruction, other types of sequences are also produced for other purposes. Some examples are:

(a) Filtered images recorded at different energy losses of the incident electrons, in order to deduce the chemical map of the specimen (see this Volume, Chap. IV Sec. 1.3), constitute an example of spectral imaging.
(b) Time series of images recorded in order to study the dynamical behavior of a specimen.
(c) Multielemental maps of the same specimen. These can be recorded

using X-ray diffraction (XDS), electron energy loss (EELS), SIMS or Auger electron (AES) spectroscopies, for instance.

(d) Very high resolution electron microscopy of crystalline specimens, recorded with the aim of deducing chemical concentration variations across an interface (subimages of crystalline units play the role of individual images).

(e) Multimodality imaging, where the same object area is analyzed using different imaging techniques.

In disregarding practical experimental problems, all these series have in common the fact that every pixel is now characterized by several values (one in each of the N images of the series). Each pixel can be described by a vector with N components, which represent either different attributes or the evolution of an attribute (as a function of time, energy, space coordinate, etc.). How do we analyze (or process) such data sets? Several groups of methods are available.

One method follows the *modelization approach*: a model is assumed (or tried) for describing the evolution of any pixel. The data for each pixel are then fitted to this model and the parameter(s) of the model deduced. Since the parameter value(s) is obtained for any pixel, new (synthesized) image(s) can be built. Parametric images are thus obtained, which concentrate the information contained within the whole experimental series, into one (or a few) image(s), that are more easily interpretable. One example of such an approach is the 'standard' method used for processing electron-energy-loss image series, in order to obtain the chemical map of an element [77, 78]. Time series can be modeled in a similar way.

Another group of methods is multivariate statistical analysis (MSA). Its aim is to extract the significant information from large data sets, while discarding redundant contributions. This is done by decomposing the total 'useful' information into a linear combination of 'basic' information components ('basic' essentially means simple and uncorrelated). For this purpose, the variance–covariance matrix ($\mathbf{Y}$) of the multidimensional data set is first built: $\mathbf{Y} = \mathbf{X}^t \cdot \mathbf{X}$, where $\mathbf{X}$ is the (centered) data matrix and $\mathbf{X}^t$ the transposed matrix. The variance–covariance matrix contains the variances of the different images (along its diagonal) and the covariances of the different image pairs. The eigenvalues and eigenvectors of this matrix are first computed. The eigenvectors associated with the decreasing eigenvalues correspond to new axes of representation. The new space of representation is orthogonal (uncorrelated components) and the number of useful components (i.e. those containing interpretable information) is generally smaller than the number of initial images. Several tools are available as aids to interpreting the factorial decomposition: the scores of images on the different factorial axes, and the scores of pixels on these axes (which can be visualized as factorial images, or eigenimages). From this first step (decomposition and interpretation), other approaches can follow:

(a) *Filtering and reconstitution.* Once the meaning of the factorial axes has been identified, an image series can be reconstituted by retaining only the 'useful' axes. This is an efficient method for performing multivariate

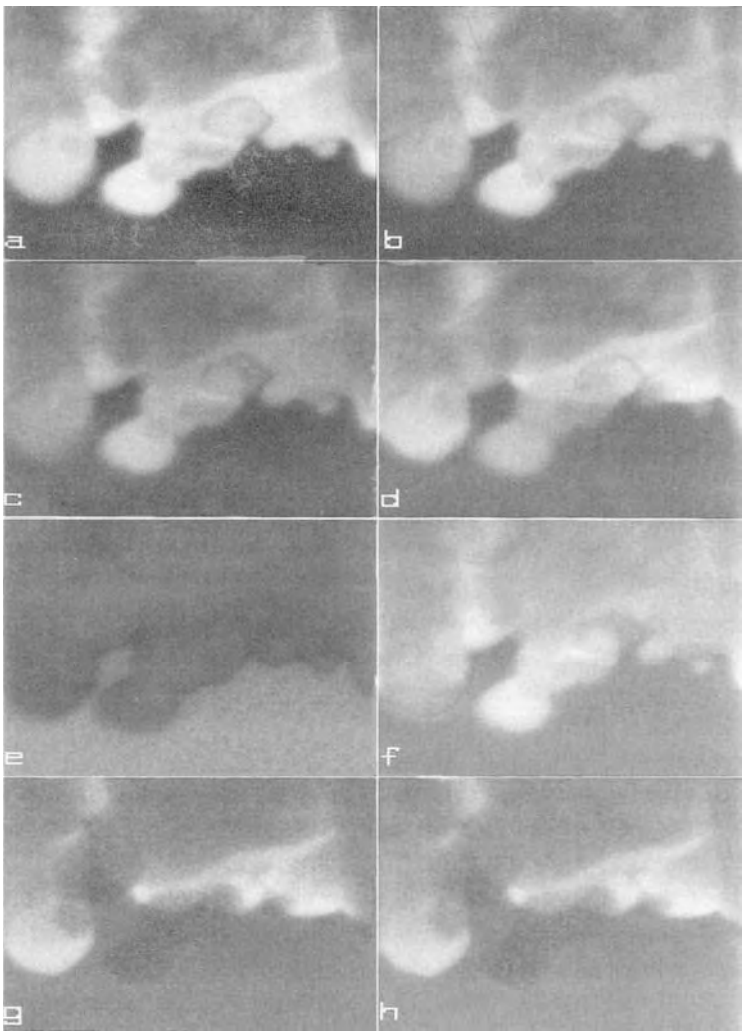


Figure 12. Illustration of different approaches for mapping chemical concentration from a sequence of electron energy filtered images (see this Volume Chap. 4 Sec. 1.3 for more details). Sequence of experimental images, recorded with a scanning transmission electron microscope and an electron-energy spectrometer at energy losses of (a) 82, (b) 99, (c) 114 and (d) 129 eV (courtesy of C. Colliex). The specimen is a Co/CeO_2 catalyst. The aim of the experiment was to map the Ce (and also Co) distribution. (g) Characteristic map of cerium. This map was obtained by subtracting the estimated background at 129 eV (f) from the experimental image at 129 eV (d). The background was estimated by extrapolating the scores of the pre-edge background images (a–c) and combining the extrapolated score with the first factorial image (e). (h) Characteristic map of cerium obtained using the standard procedure (modelization of the background according to the $A \cdot E^{-R}$ model. The similarity between the two results (g) and (h) means that the power-law model is valid in this case. But the MSA result does not assume any model. This could be an advantage in other situations. (Reproduced from [79] by permission of *Ultramicroscopy*.)

noise filtering for instance, since in the decomposition, noise is often rejected in the last factorial components.

(b) *Interpolation and extrapolation.* Instead of reconstituting the images with the initial image scores, it is also possible to reconstitute them with new coefficients. When these new values are deduced from the initial ones by interpolation or extrapolation, interpolated or extrapolated images are easily obtained [79, 80].

MSA is now increasingly applied to EDX [81–84], EELS [85–88] (see Fig. 12), Auger microanalysis [89], chemical mapping by HREM [90], and SIMS [91].

Another group of methods for the processing of image series is connected with pattern recognition. As every pixel is described by a vector of attributes, we can apply to this set of vectors the methods of pattern recognition, which consist of computing distances between vectors, angles between vectors, automatic clustering, etc. An example of this approach is the method used for quantifying the chemical-content change across an interface for HREM (lattice imaging) of crystals [92]: the image of a unit cell is represented by a multidimensional vector, the components of which are the gray levels of the different pixels that compose the unit cell (say, $30 \times 30 = 900$ components). If we assume that two vectors are available as reference vectors (those vectors which correspond to average unit cells far from the interface, for instance, where the chemical composition is known), indicators can be built which relate an unknown vector to these reference vectors. Thus, provided experimental conditions (specimen thickness and defocus) are carefully chosen so that these

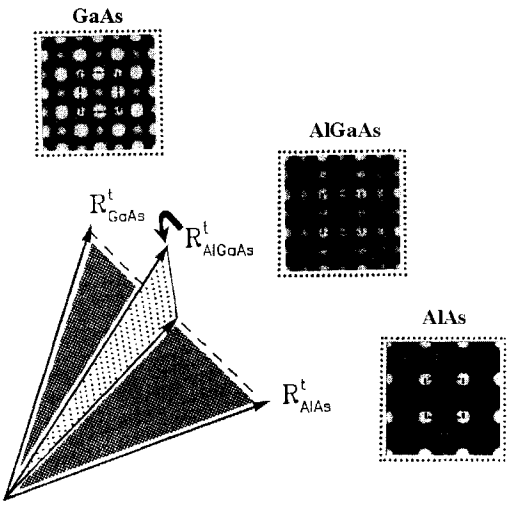


Figure 13. Principle of quantitative lattice imaging: lattice images are considered as multidimensional vectors. In general, vectors corresponding to columns with unknown composition (here $Al_xGa_{1-x}A_y$) do not lie in the plane defined by reference vectors (here GaAs and AlAs), but the projection on this plane allows the composition parameter x to be deduced. (Reproduced from [92] by permission of *Ultramicroscopy*.)

indicators can be related linearly to the concentration variation, the concentration at any point of the interface can be determined. This procedure is illustrated in Fig. 13. Recently, it has been shown that this type of method can be extended to measure the projected atomic potential, with no knowledge of the imaging conditions [93]. This approach, named QUANTITEM (quantitative analysis of the information from transmission electron micrographs) is illustrated in Fig. 14.

Another example is in the preliminary attempt to perform automatic correlation partitioning (ACP) from multielemental maps, that is to define automatically regions with homogeneous composition within a specimen for which several composition maps have been recorded. When only two maps have been recorded, the

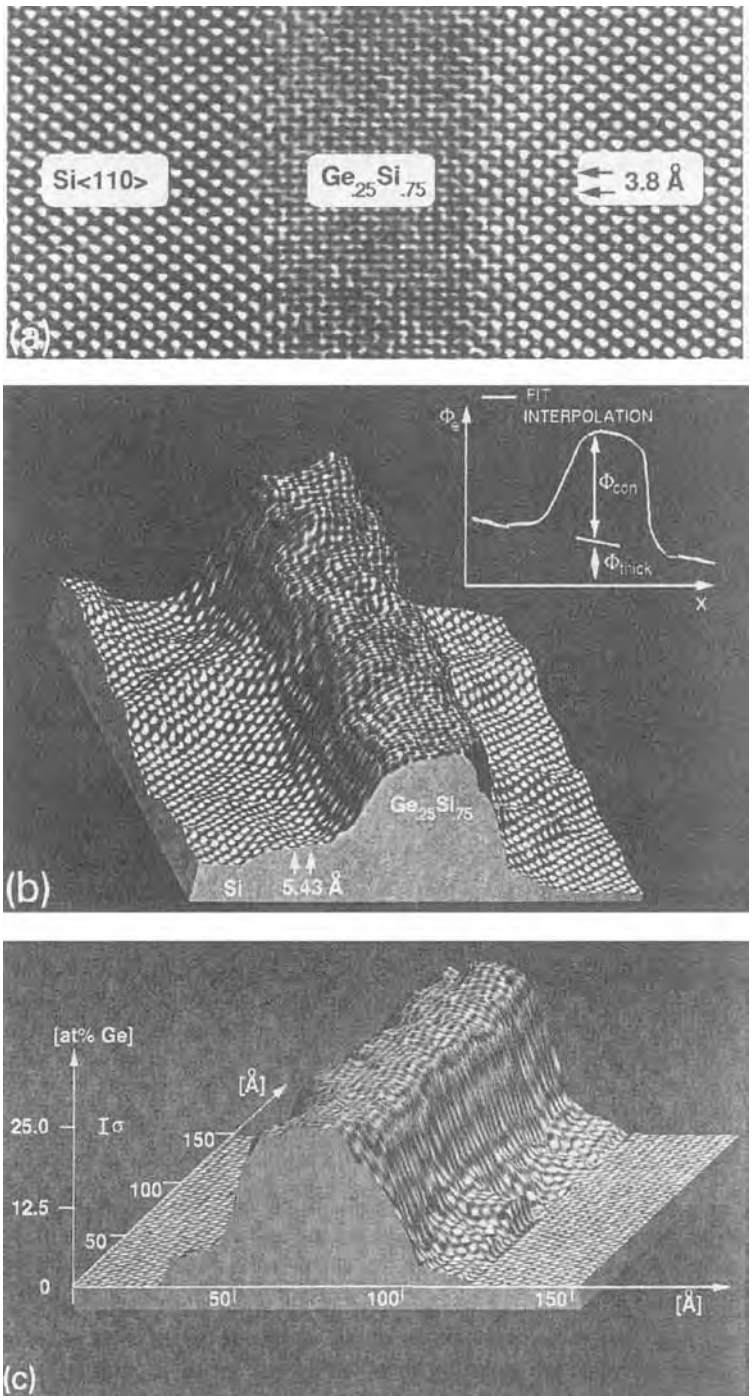


Figure 14. Illustration of the QUANTITEM procedure. (a) Experimental ⟨110⟩ lattice image of a Si/GeSi/Si structure. (b) Map of phase angle over the sample. (Inset) Schematic representation of the way in which QUANTITEM interpolates the sample thickness over regions of unknown composition, and separates the phase angle into parts according to thickness change and composition change. (c) Composition map of the Si/GeSi/Si structure (height represents composition). (Reproduced from [93] by permission of *Ultramicroscopy*.)

usual way to proceed consists of computing and displaying the scatter diagram (or scatterplot), where the content of one image is plotted as a function of the content of the other image [94–96]. The different clusters can then be selected interactively and parts of the original images (with selected properties) can be restored (interactive correlation partitioning (ICP) [97]). The purpose of ACP is to generalize the procedure to more than two maps, and towards automation. In a preliminary attempt [98], we used methods of automatic clustering (the *K*-means approach) and of image segmentation (multivariate region-growing approach) for this purpose. Here again, further developments can benefit from recent advances in the general domain of pattern recognition (fuzzy logic, etc.).

2.6 Conclusion

Image processing now plays a big part, at different levels, in the interpretation of images recorded with microscopes. The human visual system is able to perform qualitative image interpretation in most situations, but an important exception to this is high resolution microscopy, where the transfer of the information from the object wavefunction to the image intensity is very sensitive to the imaging conditions. However, when it is necessary to go from a qualitative description to a quantitative interpretation, a number of intermediate steps have to be followed, ranging from image preprocessing to image segmentation and image analysis. Moreover, one single image is often insufficient to characterize an object, and image series (focus series, tilt series, depth series, time series, energy series, etc.) are thus increasingly being produced. In this case, the need for computer algorithms to extract the useful information is even greater.

Image processing is still an evolving discipline. During the last 20 years tremendous progress has been made in different directions. New tools are becoming available every day. In addition, the theory of image processing is also being progressively improved; for example, the theory of image algebra [99–101], which is able to establish links between tasks as different as image processing by convolution, mathematical morphology, three-dimensional reconstruction and multidimensional filtering. Of course, such progress is also facilitated by the new possibilities offered by the improved capabilities of personal computers and workstations.

Perhaps the need to moderate this optimistic view about the development of image processing comes from that fact that it often takes a long time for the new possibilities, discovered by researchers in the domain of 'pure' image processing, to reach the larger community of microscopists.

2.7 References

[1] W. K. Pratt, *Digital Image Processing*, Wiley, New York, **1978**.
[2] R. C. Gonzalez, P. Wintz, *Digital Image Processing*, Addison-Wesley, New York, **1987**.
[3] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs, NJ, **1989**.
[4] J. C. Russ, *Computer-Assisted Microscopy*, Plenum Press, New York, **1990**.
[5] D. Gabor, *Lab. Invest.* **1965**, *14*, 63.

[6] A. Beghdadi, A. Le Negrate, *Computer Vision, Graphics and Image Processing* **1989,** *46*, 62.
[7] P. Perona, M. Malik, *IEEE Trans. Pattern Analysis Machine Intelligence* **1990,** *12*, 629.
[8] M. Unser, *Signal Proc.* **1990,** *20*, 3.
[9] J. Maeda, K. Murata, *Optics Commun.* **1986,** *59*, 11.
[10] C. Daly, C. Lajaunie, D. Jeulin, *Geostatistics* **1988,** *1*, 749.
[11] C. Daly, D. Jeulin, D. Benoit, *Scanning Microsc.* **1992,** Suppl. 6, 137.
[12] R. H. Bates, M. J. McDonnell, *Image Restoration and Reconstruction*, Clarendon Press, Oxford **1986**.
[13] N. Wiener, *The Extrapolation, Interpolation and Smoothing of Stationary Time Series*, Wiley, New York **1949**.
[14] S. Kawata, Y. Ichioka, *J. Opt. Soc. Am.* **1980,** *70*, 768.
[15] W. O. Saxton, *Computer Techniques for Image Processing in Electron Microscopy*, Academic Press, New York **1978**.
[16] P. W. Hawkes (Ed.), *Computer Processing of Electron Micrographs*, Springer, Berlin **1980**.
[17] P. Schiske, in *Image Processing and Computer-aided Design in Electron Optics* (Ed.: P. W. Hawkes), Academic Press, London **1973,** 82–90.
[18] W. Coene, A. Janssen, M. Op de Beeck, D. Van Dyck, *Phys. Rev. Lett.* **1991,** *69*, 37.
[19] D. Van Dyck, M. Op de Beeck, W. Coene, *Optik* **1993,** *93*, 103.
[20] A. I. Kirkland, W. O. Saxton, K. L. Chau, K. Tsuno, M. Kawasaki, *Ultramicroscopy* **1995,** *57*, 355.
[21] A. Thust, M. Lentzen, K. Urban, *Ultramicroscopy* **1994,** *53*, 101.
[22] S. J. Pennycook, D. E. Jesson, M. F. Chisholm, A. G. Ferridge, M. J. Seddon, *Scanning Microsc.* **1992,** Suppl. 6, 233.
[23] . Downing, *Scanning Microsc.* **1992,** Suppl. 6, 43.
[24] J. M. Rodenburg, B. C. McCallum, *Scanning Microsc.* **1992,** Suppl. 6, 223.
[25] D. J. Keller, F. S. Franke, *Surf. Sci.* **1993,** *294*, 409.
[26] N. Bonnet, S. Dongmo, P. Vautrot, M. Troyon, *Microsc. Microanal. Microstruct.* **1994,** *5*, 1.
[27] R. Haralick, L. Shapiro, *Scanning Microsc.* **1988,** Suppl. 2, 39.
[28] P. K. Sahoo, S. Soltani, A. K. C. Wong, Y. C. Chen, *Computer Vision, Graphics and Image Processing* **1988,** *41*, 233.
[29] C. A. Glasbey, *Computer Vision, Graphics and Image Processing* **1993,** *55*, 532.
[30] R. Deriche, *IEEE Trans. Pattern Analysis Machine Intelligence* **1990,** *12*, 78.
[31] J. Shen, S. Castan, *Computer Vision, Graphics and Image Processing: Graphical Models Image Processing* **1992,** *54*, 112.
[32] M. Petrou, *Adv. Electron. Electron Phys.* **1994,** *88*, 297.
[33] S. L. Horowitz, T. Pavlidis, *J. Assoc. Comput. Machine* **1976,** *23*, 368.
[34] S. Chen, W. Lin, C. Chen, *Computer Vision, Graphics and Image Processing: Graphical Models Image Processing* **1991,** *53*, 457.
[35] M. Kass, A. Witkin, D. Terzopoulos, *Int. J. Comput. Vis.* **1988,** *1*, 321.
[36] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, London **1982**.
[37] M. Coster, J. L. Chermant, *Précis d'Analyse d'Images*, CNRS, Paris **1989**.
[38] E. R. Dougherty, *Mathematical Morphology in Image Processing*, Dekker, New York **1992**.
[39] D. Jeulin, *Scanning Microsc.* **1988,** Suppl. 2, 165.
[40] S. Beucher, *Scanning Microsc.* **1992,** Suppl. 6, 299.
[41] S. Beucher, F. Meyer, in *Mathematical Morphology in Image Processing* (Ed.: E. R. Dougherty) Dekker, New York, **1992** 433.
[42] S. O. Belkasim, M. Shridhar, M. Ahmadi, *Patt. Rec.* **1991,** *24*, 1117.
[43] L. M. Luerkens, J. K. Beddow, A. F. Vetter, *Powder Technol.* **1982,** *31*, 209.
[44] L. M. Karlsson, A. Liljeborg, *J. Microsc.* **1994,** *175*, 186.
[45] J. C. Russ, D. S. Bright, J. C. Russ, T. M. Hare, *J. Comput. Assist. Microsc.* **1989,** *1*, 3.
[46] E. R. Davies, *Machine Vision*, Academic Press, London **1990**.
[47] J. Frank, in *Computer Processing of Electron Micrographs* (Ed.: P. W. Hawkes), Springer, Berlin **1980,** 187.
[48] R. O. Duda, P. E. Hart, *Pattern Recognition and Scene Analysis*, Wiley, New York **1973**.
[49] R. M. Haralick, *Proc. IEEE* **1979,** *67*, 786.
[50] K. Laws, *Textured Image Segmentation*, Report 940, USC Image Processing Institute, Los Angeles, CA **1980**.
[51] M. Unser, M. Eden, *IEEE Trans. Syst. Man Cybern.* **1990,** *20*, 804.
[52] A. K. Jain, F. Farrokhnia, *Patt. Rec.* **1991,** *24*, 1167.
[53] B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco **1982**.
[54] M. Aguilar, E. Anguiano, F. Vasquez, M. Pancorbo, *J. Microsc.* **1992,** *167*, 197.
[55] A. P. Pentland, *IEEE Trans. PAMI* **1984,** *6*, 661.
[56] J. C. Russ, *J. Comput. Assist. Microsc.* **1990,** *2*, 249.

[57] S. Peleg, J. Naor, R. Hartley, D. Avnir, *IEEE Trans. Pattern Analysis Machine Intelligence* **1984,** *6*, 518.
[58] E. E. Underwood, *Quantitative Stereology*, Addison-Wesley, Reading, **1970**.
[59] J. C. Russ, *Practical Stereology*, Plenum Press, New York **1986**.
[60] L. M. Cruz-Orive, *Acta Stereol.* **1987,** *6*, 3.
[61] H. J. Gundersen, *Acta Stereol.* **1987,** *6*, 173.
[62] M. Prod'homme, L. Chermant, M. Coster, *J. Microsc.* **1992,** *168*, 15.
[63] M. Prod'homme, M. Coster, L. Chermant, J. L. Chermant, *Scanning Microsc.* **1992,** Suppl. 6, 255.
[64] D. Jeulin, *Signal Proc.* **1989,** *16*, 403.
[65] D. Jeulin, *Scanning Microsc.* **1992,** Suppl. 6, 121.
[66] G. Matheron, *Random Sets and Integral Geometry*, Wiley, New York **1975**.
[67] E. Pirard, *J. Microsc.* **1994,** *175*, 214.
[68] S. K. Gosh, *Methods Cell Biol.* **1981,** *22*, 155.
[69] I. Stoev, S. Simov, E. Simova, N. Bonnet, G. Balossier, *Sensors Actuators* **1987,** *12*, 1.
[70] D. J. Fleet, A. D. Jepson, R. M. Jenkin, *Computer Vision, Graphics and Image Processing: Image Understanding* **1991,** *53*, 198.
[71] J. Weng, *Int. J. Comput. Vision* **1993,** *11*, 211.
[72] E. Zeitler, *Optik* **1974,** *39*, 396.
[73] G. T. Herman, *Image Reconstruction from Projections: the Fundamentals of Computerized Tomography*, Academic Press, New York **1980**.
[74] J. Frank (Ed.), *Three-Dimensional Imaging with the Transmission Electron Microscope*, Plenum Press, New York **1992**.
[75] K. H. Downing, H. Meisheng, H.-R. Wenk, M. O'Keefe, *Nature* **1990,** *348*, 525.
[76] L. Lucas, N. Gilbert, D. Ploton, N. Bonnet, *J. Microsc.* **1996,** *181*, 238.
[77] C. Jeanguillaume, C. Colliex, P. Trebbia, *Ultramicroscopy* **1978,** *3*, 137.
[78] N. Bonnet, C. Colliex, C. Mory, M. Tence, *Scanning Microsc.* **1988,** Suppl. 2, 351.
[79] N. Bonnet, E. Simova, S. Lebonvallet, H. Kaplan, *Ultramicroscopy* **1992,** *40*, 1.
[80] N. Bonnet, P. Trebbia, *Scanning Microsc.* **1992,** Suppl. 6, 163.
[81] R. Browning, P. King, J. M. Paque, P. Pianetta, *Microbeam Anal.* **1990,** 199.
[82] C. Quintana, N. Bonnet, *Scanning Microsc.* **1994,** *8*, 563.
[83] C. Quintana, N. Bonnet, *Scanning Microsc.* **1994,** Suppl. 8, 83.
[84] P. Trebbia, J. M. Wulveryck, N. Bonnet, *Microbeam Anal.* **1995,** *4*, 85.
[85] P. Hannequin, N. Bonnet, *Optik* **1988,** *81*, 6.
[86] P. Trebbia, N. Bonnet, *Ultramicroscopy* **1990,** *34*, 165.
[87] P. Trebbia, C. Mory, *Ultramicroscopy* **1990,** *34*, 237.
[88] E. Gelsema, A. Beckers, W. De Bruijn, *J. Microsc.* **1994,** *174*, 161.
[89] M. Prutton, M. El Gomati, P. Kenny, *J. Electron Spectrosc. Relat. Phenom.* **1990,** *52*, 197.
[90] J. L. Rouvière, N. Bonnet, *Inst. Phys. Conf. Ser.* **1993,** *134*, 11.
[91] P. Van Espen, G. Janssens, W. Vanhoolst, P. Geladi, *Analysis* **1992,** *20*, 81.
[92] A. Ourmazd, F. H. Baumann, M. Bode, Y. Kim, *Ultramicroscopy* **1990,** *34*, 237.
[93] C. Kisielowski, P. Schwander, F. H. Baumann, M. Seibt, Y. Kim, A. Ourmazd, *Ultramicroscopy* **1995,** *58*, 131.
[94] C. Jeanguillaume, *J. Microsc. Spectrosc. Electron.* **1985,** *10*, 409.
[95] D. Bright, D. Newbury, *Anal. Chem.* **1991,** *63*, 243.
[96] P. Kenny, I. Barkshire, M. Prutton, *Ultramicroscopy* **1994,** *56*, 289.
[97] J. M. Paque, R. Browning, P. L. King, P. Pianetta, *Microbeam Anal.* **1990,** 195.
[98] N. Bonnet, *Ultramicroscopy* **1995,** *57*, 17.
[99] G. Y. Ritter, *Adv. Electron. Electron Phys.* **1991,** *80*, 243.
[100] P. W. Hawkes, *Optik* **1993,** *93*, 149.
[101] P. W. Hawkes, *Scanning Microsc.* **1994,** Suppl. 8, 289.

Part IX

Special Topics

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

1 Coincidence Microscopy

1.1 Introduction

The impact of a single fast electron on a specimen in the electron microscope can cause several events: impact ionization accompanied by emission of an electron, subsequent decay of the excited state accompanied by the emission of an X-ray or an Auger-electron and, of course, energy loss of the primary electron. All these events may occur in combination with elastic scattering, backscattering, other secondary processes, or even absorption. The time scale on which these events follow each other is usually in the 10^{-14} s range and hardly ever exceeds 10^{-9} s. Since it is technically feasible to detect these events with a time resolution of the order of 10^{-9} s, there is a possibility of obtaining more than one signal from a single scattering event. Such a measurement, involving two signals that coincide in time, is called a coincidence measurement. It contains more information than the two signals involved can give separately. The best way to think about this is to consider one signal as a filter for the spectrum of the other signal: this spectrum can be split up in the different parts that constitute it in the first place. For example, in the energy loss signal it is possible to select a specific excitation and then, in the coincidence–Auger spectrum, measure the decay products of only that excitation. Or, the one signal can be secondary electrons and then, in the coincidence–energy loss spectrum, it is possible to find only the events that were responsible for the emission of a secondary electron.

Historically, the coincidence technique was developed in the field of elementary particle physics, but the power was soon recognized in atomic and molecular physics. The most common application is to determine the electron momentum distribution, for which a relatively low energy electron beam scatters on a gas and both the energy and angular direction of the two outgoing electrons is selected [1]. More recently, these (e,2e) experiments have also been performed on solids, at higher energies, but not yet with a focussed primary beam [2–6]. VanderWiel and Brion [7] have reported on near-dipole (e,2e) spectroscopy, which uses the similarity between photon and electron-induced excitations. By acquiring the Auger electrons in coincidence with certain well-defined energy losses in a gas, Ungier and Thomas [8] could decompose Auger spectra to study specific decay processes. A different form of coincidence spectroscopy measures the photoelectron and the Auger electron after absorption of an X-ray [9, 10]. Coincidence measurements in the

electron microscope have been reported [11–17], but hardly using the high spatial resolution of the microscope and not yet applied to obtain information on the specimen. After these proofs of principle, the method awaits improvements in instrumentation. One potential improvement is the development of parallel detectors with the required timing resolution; another improvement would be a more generally available high vacuum, since the most promising form of coincidence measurement involves spectroscopy of emitted electrons.

1.2 Instrumentation

For coincidence microscopy, an STEM is the most appropriate instrument, although it is possible to use a TEM with a small probe or, for some forms of coincidence measurement, an SEM. The two signals that are to be measured in coincidence must be detected with a high detection efficiency, since coincidence rates are usually low. More specific for the coincidence measurements, however, is that single events must be detected with a good time resolution. For the signal processing, some dedicated electronics is necessary.

Intrinsic to electron energy loss spectroscopy (EELS) in the STEM is the very high collection efficiency, because the fast primary electrons are scattered within a small solid angle. In fact, this is the main reason for an electron microscope being the ideal instrument for coincidence measurements. If the system has a serial acquisition mode, there usually is a slit behind the spectrometer and a scintillator–photomultiplier tube (PMT) for detection. If the scintillator has a short decay time, as with plastic scintillators although not the case with some other materials, it is possible to measure single event pulses at the exit of the PMT. These can be used for the coincidence measurement. A parallel detector behind the spectrometer is usually a photodiode array or a CCD camera. In such devices all timing information is lost. Parallel detectors for high energy electrons with sufficient timing resolution and the ability to detect up to 10^6 or 10^7 electrons per second are expected to become available in the near future.

Energy dispersive X-ray detectors are single event sensitive and are parallel in the sense that the energy of each X-ray photon is determined. However, the time resolution is very limited due to the small number of electron hole pairs that are created in the pn-junction. The time necessary to determine the X-ray energy is several hundreds of nanoseconds although, using a second signal path, the timing can be improved to several tens of nanoseconds [13]. For any further improvement it would be necessary to use a wavelength dispersive spectrometer with a position sensitive detector. The disadvantage of WDS detectors, however, lies in their lower collection acceptance angle.

Secondary electron detectors using fast scintillator PMT combinations and cathodoluminescence detectors using PMTs are well suited for coincidence measurements.

Energy spectrometers for secondary and Auger electrons in a microAuger instrument usually have single electron detectability, although the collection efficiency is relatively poor and there is no EELS signal, unless the experiment is performed in reflection [18]. The Auger spectrometers especially designed for

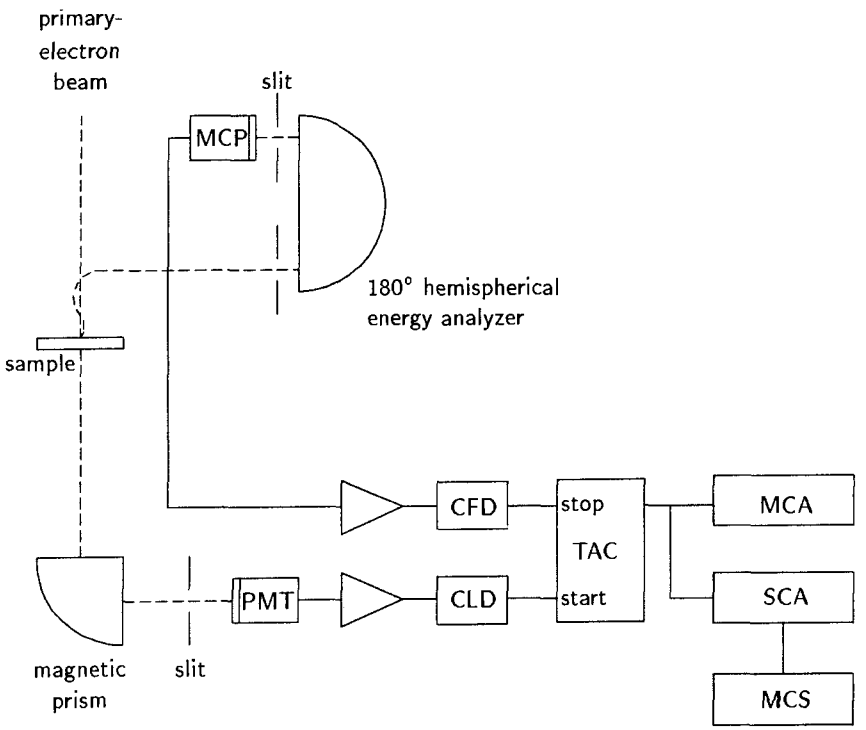


Figure 1. Experimental setup for EELS–emitted electron coincidence measurements. Key: (MCP) microchannel plate; (PMT) photomultiplier tube; (CFD) constant fraction discriminator; (CLD) constant level discriminator; (TAC) time-to-amplitude converter; (MCA) multichannel analyzer; (SCA) single channel analyzer; (MCS) multichannel scaler.

STEM instruments [19] combine a high collection efficiency and a high spatial resolution. The secondary and Auger electrons spiral up through the magnetic field of the objective lens pole piece and are subsequently deflected to enter a spectrometer. Behind the spectrometer, the electrons are detected with a channeltron or multichannelplate, giving excellent time resolution. The multichannelplate can be combined with a position-sensitive detector for parallel detection, either based on the resistive anode techique or with multiple anodes. The latter can accept a higher count rate.

The electronics necessary for serial–serial coincidence measurements are schematically shown in Fig. 1. Pulses from the detectors are amplified and fed into discriminators to obtain TTL or ECL level pulses. The best timing resolution is obtained when using constant fraction discriminators, which trigger the output pulse not at a predefined level, but rather at a level which is defined as a fraction of the amplitude of the input pulse. A time-to-amplitude converter, or a time-to-digital converter, creates a signal proportional to the time interval between the arrival of pulses from the two detectors. A multichannel scaler can form a time spectrum in which it is possible to find the coincidence peak on top of a false coincidence background. Setting a

window on the time axis, the events that follow within the window can be counted as a function of the energies selected in the spectrometers. This yields the coincidence spectrum. The use of parallel detectors involves a more complicated setup of the electronics and data processing [20].

1.3 Coincidence Count Rates

Given a primary beam current I and the cross section for a specific event α to occur σ_α, the count rate connected with this event is

$$R_\alpha = \frac{I}{e}\,\sigma_\alpha n t f_\alpha T_\alpha \eta_\alpha \tag{1}$$

where e is the electron charge, n is the atomic density, t the thickness of the specimen, f_α the fraction of the thickness from which information can escape and contribute to the signal, T_α the transmission efficiency to the detector and η_α the efficiency of the detector. Alternatively, the countrate can be expressed in terms of escape depth when this is smaller than the thickness of the specimen [10]. If the same primary electron can also cause an event β, a similar equation holds for the count rate R_β. For example, α can be the K shell excitation of carbon, with an energy loss between 300 and 305 eV, β can be the Auger decay emitting an electron of energy between 260 and 265 eV. If both events are measured in coincidence, the coincidence count rate is

$$R_\mathrm{T} = \frac{I}{e}\,\sigma_\mathrm{true} n t f_\alpha f_\beta T_\alpha T_\beta \eta_\alpha \eta_\beta \tag{2}$$

where σ_true is the cross section for the combined event α and β. There is a possibility that two unrelated events in the detectors occur at the same time, or more precisely, within the time window τ which is set to define which events are coincident. The associated count rate, called false coincidence rate, is given by

$$R_\mathrm{F} = R_\alpha R_\beta \tau \tag{3}$$

From the ratio of Eq. (2) to Eq. (3), and substituting for R_α and R_β using Eq. (1), we find the true to false ratio

$$\frac{R_\mathrm{T}}{R_\mathrm{F}} = \frac{1}{nt} \times \frac{\sigma_\mathrm{true}}{\sigma_\alpha \sigma_\beta} \times \frac{e}{I} \times \frac{1}{\tau} \tag{4}$$

This shows how important it is to obtain a good time resolution so that τ can be made small. Although it seems advantageous to also make the beamcurrent small, this is somewhat misleading: the false coincidence background can easily be subtracted, so it becomes a question of statistical noise in the background subtracted spectra. From that, an optimized current follows for each experiment [21]. The numerical value of the count rate depends very much on the process under observation and on the instrumental setup. For EELS–secondary electron measurements at the plasmon excitation, with energy windows of 1 eV in both spectra, coincidence rates of the order of 1000 counts per second were obtained for a primary beam current less than 1 nA. The false coincidence background was an order of magnitude smaller.

1.4 Signal Combinations

1.4.1 EELS–Emitted Electron

Since the energy distributions of both the EELS and the emitted electrons contain

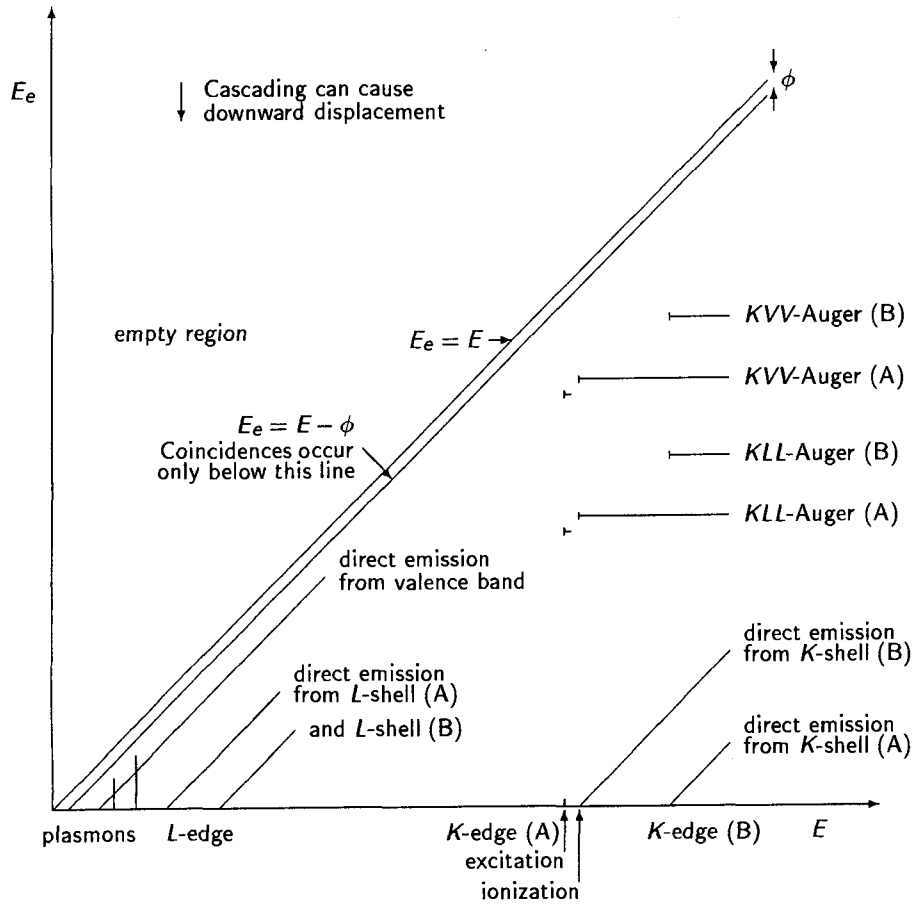


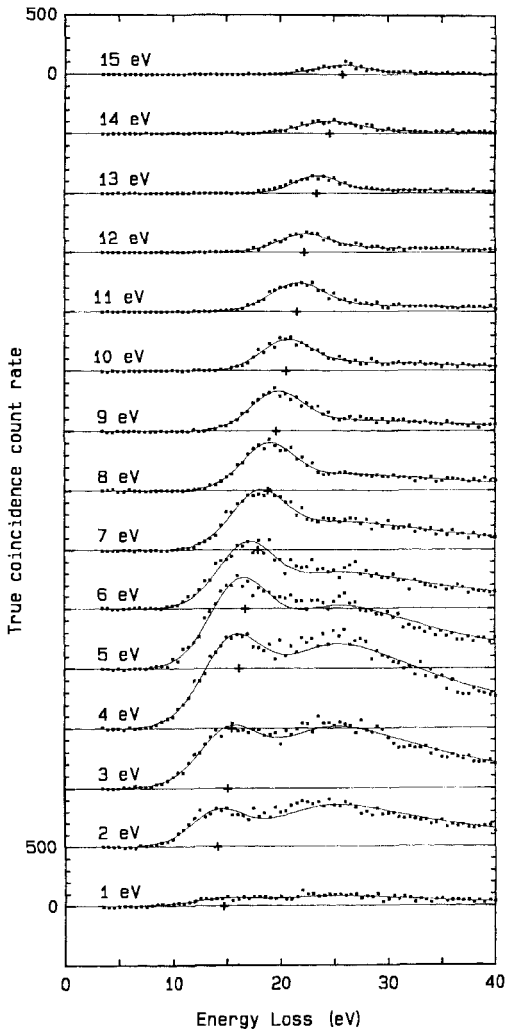
Figure 2. Schematic 2D EELS–emitted electron coincidence spectrum, for a specimen with two elements, A and B. Along the horizontal axis is the energy loss of the transmitted electron, along the vertical axis is the kinetic energy of the emitted electron. (From Pijper [21], reproduced by permission of the author.)

information, a coincidence measurement between these two signals is a two-dimensional spectrum. Figure 2 gives a schematic representation of such a spectrum, which can be used to discuss some of the applications of this signal combination.

The low energy part of the EELS spectrum gives information on plasmon excitation and single electron excitation from the valence band. If counts in the EELS spectrum are only accepted if in coincidence with a secondary electron, one effectively obtains the energy loss events which are responsible for emitting a secondary electron. Voreades [11] first performed this experiment although without energy analysis of the secondary electrons. His conclusion was that most secondary electrons are produced by energy loss events of about 20 eV, but it was impossible to decide whether plasmon decay processes or interband transitions were predominantly responsible for the secondary electrons. Pijper and Kruit [14] worked with a

better EELS resolution and were able to discriminate for different secondary electron energies, thus, they could identify the surface plasmon excitation and decay as an important contributor to secondary electrons in the case of amorphous carbon films. For all other energy loss processes, including plasmon excitation, they established the validity of the stopping power rule: the probability of emitting a secondary electron is directly proportional to the amount of energy lost by the primary electron. In addition, it was clear from the spectra, of which an example is shown in Fig. 3, that direct excitation from the valence band can be seen as a feature in the 2D-spectrum that runs diagonally from the bottom-left to the top-right: the energy of the emitted electron is directly proportional to the energy lost by the primary electron. Similar experiments with EELS–secondary electron coincidence were performed by Mullejans et al. [15], and Scheinfein et al. [16]. The latter showed that in Si, it is not the plasmon which is responsible for most of the secondary electrons. Drucker et al. [22] extended the technique by also discriminating energy loss events on the basis of scattering angle and concluded that secondary electrons are more efficiently produced by large momentum transfer scattering, thus explaining the very high spatial resolution obtainable in secondary electron imaging.

Perhaps the most promising application of coincidence microscopy is virtual photoelectron spectroscopy. Because of the similarity between forward electron scattering and photoabsorption, a vertical section through the 2D coincidence spectrum resembles a photoelectron spectrum. Selecting a different energy loss at which



4 4

Figure 3. Energy loss coincidence spectra (shown as dots) for 15 consecutive values of the secondary electron energy, 1 eV apart. The spectra have been vertically displaced over an equivalent of 500 counts. False coincidence counts have been subtracted. Solid curves are fitted to the data with a model including a plasmon function and a Gaussian. The position of the Gaussian peak is indicated with a +.

the section is taken is equivalent with tuning to a different photon energy in a synchrotron. The positions of the photoelectron peaks shift with the selected energy loss value, indicated by the diagonal lines in Fig. 2. The advantage over photoelectron spectroscopy lies in the obtainable spatial resolution. The energy resolution is limited by the EELS resolution. Whether spatially resolved virtual-photoelectron spectroscopy will indeed become a useful technique will depend on the possibility of obtaining a parallel detection coincident spectra in an ultra-high vacuum electron microscope. It would then be relatively easy to extend the technique to select specific scattering angles of the primary electron. However, to determine also the angular distribution of the emitted electrons, in order to have a full (e,2e) experiment, seems to be much more complicated.

A normal Auger spectrum is the vertical spectrum that is obtained after integration over all possible energy losses in Fig. 2. A coincident Auger spectrum is the spectrum obtained after integration only over a selected range of energy losses. If only those energy loss events are selected which have a relatively high probability of producing an Auger electron, an improvement of peak-to-background ratio in the Auger spectrum can be expected, comparable to the improvement obtained in X-ray excited Auger spectra. This might be the most promising approach to single atom identification, since it combines a high signal-to-background spectrum with an extremely small analyzed volume. Another application of the coincidence Auger spectroscopy might be to study the decay of weakly bound states near the absorption edge in the EELS spectrum, equivalent to the experiments of Ungier and Thomas [8] on gases. For solids, this would yield information on local chemical bonds.

Quite similar to the coincident Auger spectrum, a coincident EELS spectrum is obtained after integration over a selected range of Auger energies in Fig. 3. Again, the effect is background reduction, but now accompanied with an appreciable loss of signal, because Auger electrons are only detected if they come from the surface of the specimen.

1.4.2 EELS–X-ray

The first effort using coincidence techniques in the electron microscope that was directed at improving the detection limits of microanalysis used the coincidence between X-rays and energy loss events [13]. An energy loss spectrum around the Ca K-edge collected in coincidence with CaK X-rays, showed an almost complete removal of the background under the edge. A background reduction could also be obtained in the X-ray spectrum. However, due to the relatively slow EDX detector, the primary current had to be low and thus there was no improvement of the signal-to-noise ratio. Attempts by Nicholls et al. [23] were no more successful.

1.4.3 EELS–Cathodoluminescence

By measuring the time delay between an electron energy loss event and the corresponding cathodoluminescence event, lifetimes of excited states may be determined

[12, 24]. As in the case of secondary electron production, the CL production efficiency is linearly proportional to the energy loss, except at low energies. The technique has been applied to scintillators and semiconductors. The originators of the method mention the possibility of producing submicron resolution maps of lifetimes in the vicinity of inhomogeneities in semiconductor crystals. Selecting only short lived excitations might improve the resolution in cathodoluminescence images.

1.4.4 Backscattered Electron–Secondary Electron

For bulk samples, the energy loss signal must be obtained in reflection mode, or be accompanied by a backscatter event. Preliminary experiments have been performed by Kirschner et al. [17] and Cazaux et al. [18]. Further experiments are proposed by Gergely [25].

1.4.5 Other Combinations

Other signal combinations are possible. For example, accepting X-rays only when they are in coincidence with secondary electrons would give surface sensitivity to the X-ray microanalysis technique. Many energy loss events give rise to more than one emitted electron, or to one electron plus a photon or X-ray. Technically it is possible to perform triple coincidence measurements, which should lead to results comparable with Auger-photoelectron coincidence experiments [9, 10].

1.5 References

[1] I. E. McCarthy, E. Weigold, *Rep. Prog. Phys.* **1991,** *54*, 789.
[2] U. Amaldi Jr., A. Egidi, R. Marconero, G. Pizzella, *Rev. Sci. Instrum.* **1969,** *40*, 1001.
[3] R. Camilloni, A. Giardini Guidoni, R. Tiribelli, G. Stefani, *Phys. Rev. Lett.* **1972,** *29*, 618.
[4] N. A. Krasil'nikova, N. M. Persiantseva, *Phys. Lett.* **1978,** *69A*, 287.
[5] A. L. Ritter, J. R. Dennison, R. Jones, *Phys. Rev. Lett.* **1984,** *53*, 2054.
[6] Chao Gao, A. L. Ritter, J. R. Dennison, N. A. W. Holzwarth, *Phys. Rev. B* **1988,** *54*, 3914.
[7] M. J. van der Wiel, Th. M. El-Sherbini, C. E. Brion, *Chem. Phys. Lett.* **1970,** *7*, 161.
[8] L. Ungier, T. D. Thomas, *J. Chem. Phys.* **1985,** *82*, 3146.
[9] H. W. Haak, G. A. Sawatsky, T. D. Thomas, *Phys. Rev. Lett.* **1978,** *41*, 1825.
[10] E. Jensen, R. A. Bartynski, S. L. Hulbert, E. D. Johnson, *Rev. Sci. Instrum.* **1992,** *63*, 3013.
[11] D. Voreades, *Surf. Sci.* **1976,** *60*, 325.
[12] R. J. Graham, J. C. H. Spence, H. Alexander, *Mat. Res. Soc. Symp. Proc.* **1987,** *82*, 235.
[13] P. Kruit, H. Shuman, A. P. Somlyo, *Ultramicroscopy* **1984,** *13*, 205.
[14] F. J. Pijper, P. Kruit, *Phys. Rev. B* **1991,** *44*, 9192.
[15] H. Mullejans, A. L. Bleloch, *Phys. Rev. B* **1992,** *46*, 8597.
[16] M. R. Scheinfein, J. Drucker, J. K. Weiss, *Phys. Rev. B* **1993,** *47*, 4068.
[17] J. Kirschner, O. M. Artomonov, A. N. Trekhov, *Phys. Rev. Lett.* **1992,** *69*, 1711.
[18] J. Cazaux, O. Jbara, K. H. Kim, *Surf. Sci.* **1991,** *247*, 360.
[19] P. Kruit, J. A. Venables, *Ultramicroscopy* **1988,** *25*, 183.
[20] R. W. Hollander, V. R. Bom, C. W. E. van Eijk, J. S. Faber, H. Hoevers, P. Kruit, *Nucl. Instrum. Methods A* **1994,** *348*, 664.
[21] F. J. Pijper, PhD Thesis, Delft University of Technology **1993,** p. 97.
[22] J. Drucker, M. R. Scheinfein, *Phys. Rev. B* **1993,** *47*, 15973.
[23] A. W. Nicholls, I. P. Jones, M. H. Loretto, *IOP Conf. Proc.* (EMAG'85), **1985,** *78*, 205.
[24] C. C. Ahn, O. L. Krivanek, *Proc. 43rd EMSA*, San Francisco Press **1985,** p. 406.
[25] G. Gergely, *Acta Physica Hungarica* **1991,** *70*, 47.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

2 Low Energy Electron Holography and Point-Projection Microscopy

2.1 Introduction and History

Electron holography has its origins in the first observations of Fresnel fringes using electrons in 1940, and with shadow imaging by projection from a small source.

Shadow images cast on a wall by an opaque object placed in front of a lamp have their exact parallel in electron microscopy – we see them in the central disc of every convergent-beam microdiffraction pattern if the probe is focused slightly above the sample at z_1, as shown in Fig. 1a. If the source is small, Fresnel fringes will be seen around object edges (indicating that the image is out of focus by z_1), and the magnification is approximately $M = z_2/z_1$, as shown. It is therefore not surprising that this mode was one of the first to be used to obtain electron microscope images in 1939 [1]. At that time, because of the large sources used, the troublesome Fresnel edge fringes were not seen. A year later, Fresnel fringes were observed for the first time using electrons by Boersch [2] (using the point-projection geometry), and by Hillier [3] (using the more familiar geometry shown in Fig. 1b). Boersch's remarkable pre-war paper gives the relationship between source size and fringe spacing. Hillier interpreted hi result in terms of Fresnel fringes an spherical aberration. In the recent light c De Broglie's matter-wave hypothesis an Davison, Germer and Thomson's worl the significance of these results was see as a confirmation of the wave nature of th electron. Six years after that, an entirel new interpretation was given by Gabor t Boersch's Fresnel edge fringes [4] when h named them in-line electron hologram and pointed out that these fringes migl be removed by a focus-restoration proc dure involving optical reconstruction. I crudest terms, it was planned to sharpe the edge of a shadow image in order t improve resolution. The shadow image the hologram (if the source is small), an reconstruction removes the Fresnel edg fringes. For these mask-like opaqu objects, the unobstructed wave passin around the edge provides a limited refe ence wave. Gabor extended the analysis t include partially transparent objects, an the aim of his work was to use in-lin holography to eliminate the aberration of the electron lens (shown in gray i Fig. 1a) used to focus the probe to 'point' source. Three-dimensional recor structions and off-axis optical holograph came much later. The first experimenta

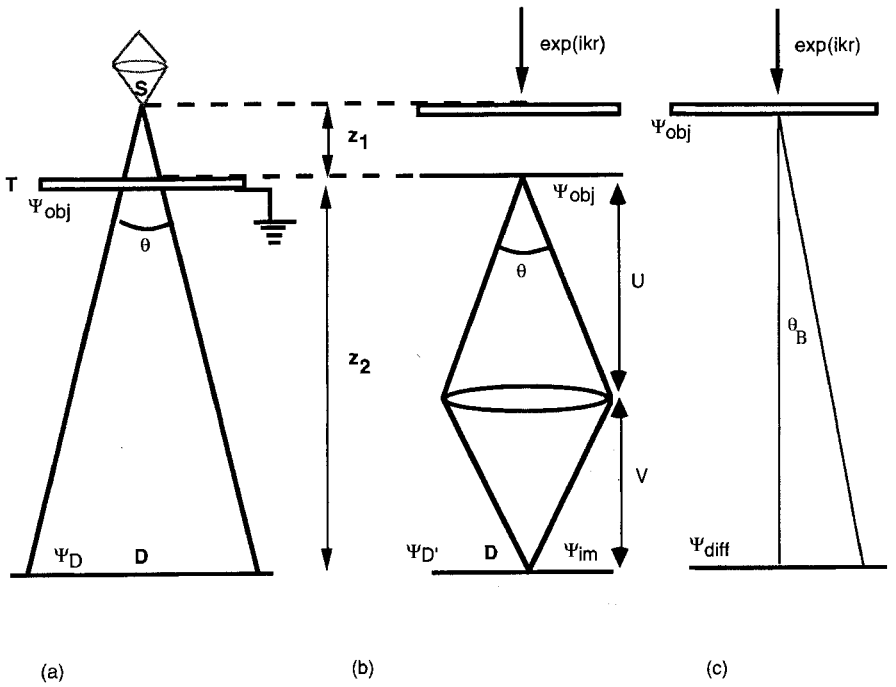


Figure 1. (a) Point-projection geometry in convergent-beam microdiffraction. The probe is focus distance z_1 slightly above the sample. (b) Transmission geometry with focus error z_1 $(1/f = 1/U + 1$, Diffraction pattern.

test of Gabor's ideas came in 1952 when Haine and Mulvey [5] compared the 'projection' (Fig. 1a) and 'transmission' (Fig. 1b) geometries, the equivalence of which was not fully appreciated at that time [6]. The success of this approach (usually using the 'transmission' geometry) has had to wait more than 40 years for the commercial availability of brighter electron sources and better lenses, and is described in detail in Sec. 2.2 of Chap. 4, this volume. Section 1.9 of Chap. 4, this Volume describes high-energy electron holography experiments in the projection geometry, using scanning transmission electron microscopy (STEM).

The appeal of the point-projection geometry lies in the fact that, if a sufficiently small source can be found, almost unlimited magnification can be ob apparently without aberrations. The rations of a probe-forming magnet can be avoided by using a physical el emitter as the source, and the first e mental shadow images formed a energy by this method appear to be of Morton and Ramberg [7] in 1939. researchers used a glass ultra-high va system and electron field-emitter close to a transmission electron scope (TEM) grid, with observatio distant screen to obtain a shadow i This approach was later develop Melmed and co-workers for both el and ion point-projection microsco Since any electric field distributio cylindrical symmetry forms a len field-emission tip itself forms a len

emitting area on the surface of the tip is imaged onto a virtual source inside the tip. The aberration coefficients of such a nano-tip lens are finite, and will be discussed below. The point-projection geometry was proposed again for the atomic-resolution electron imaging of thin crystals by Cowley and Moodie in 1957 [9] with their development of the theory of Fourier imaging. Finally, the development of field-emission STEM instruments [10] has allowed the continuous development of in-line electron holography in the point-projection geometry shown in Fig. 1a, using one stationary probe position for the collection of each hologram [11–13].

Claims for atomic resolution at low voltage using the lensless, field-emission arrangement of Morton and Ramberg have recently revived interest in this technique [14]. Using specially prepared 'single atom' field emission tips [15], Fink and co-workers have obtained point-projection images of greatly improved quality [16], showing many Fresnel edge fringes. Since the reconstructed-image resolution is approximately equal to the source size, an atomic-sized emitter should produce atomic-resolution images; however, unless there is clear evidence of penetration of the beam through the sample, the reconstructed images will show only the outline of the object (mask reconstruction). By using Fraunhofer reconstruction methods we can test for transmission of the beam through the sample. Several similar point-projection microscopes (PPM) have since been constructed by other groups [17–19]. This low-voltage approach solves the problems of lack of source brightness and lens aberrations which limited Haine and Mulvey's work; however, the severely limited penetration of low voltage electrons through matter is a severe constraint. Solutions to this problem using either the reflection geometry or very low or higher voltage are discussed later in this section.

In a typical point-projection low voltage microscope [16], a tungsten field emitter at a potential of about -100 V is placed a distance z_1 (a few thousand Angstrom) from a grounded sample, which acts as the anode. A shadow image is formed on a channel plate a distance of $z_2 = 14$ cm away, with a magnification $M = z_2/z_1$. Images of the fiber-optic face-plate may be recorded using a liquid-nitrogen-cooled CCD camera. Typical exposure conditions are about 0.1 s with less than 1 nA beam current at 100 V. The image (see Fig. 1) is superimposed on a point-projection field electron image of the tip (the tip emission pattern, projected from the virtual source inside the tip). This causes a slowly varying background modulation. Note that only movement of the tip relative to the sample is magnified; as in field ion microscopy (FIM), movement of the tip does not cause magnified movement of the field-emission pattern from the tip, since no movement of the virtual source relative to the surface of the tip occurs.

If the tip (and detector) are modeled as coaxial parabolae, the on-axis electric field $E(z)$ is given by

$$E(z) = \frac{U_0}{z \ln(2z_1/r)} \approx \frac{U_0}{6z}$$

if $z_1 = 200$ nm and $r = 1$ nm, where r is the tip radius, z is a coordinate along the axis from the focus and U_0 is the beam potential. The field varies slowly with z_1 but rapidly with r, and falls off very rapidly away from the tip. The virtual source size $d \approx r/2$ [20]. Field emission from tungsten

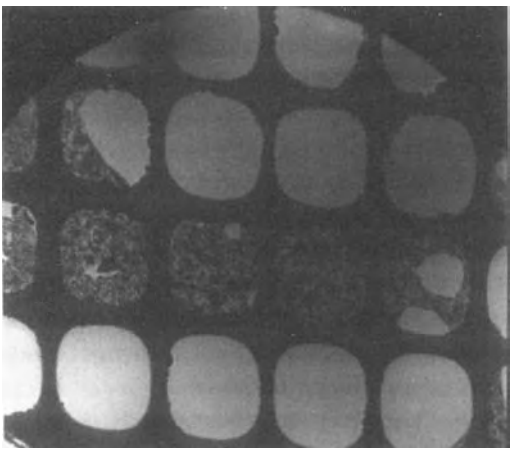


Figure 2. Low magnification image of a porous carbon film obtained at ca 400 V.

requires about $0.4\,\mathrm{V\,A^{-1}}$ at the surface, so fixing $z = r/2$ fixes $U_0 \approx 12\,\mathrm{V}$. (A more realistic value would be $U_0 = 100\,\mathrm{V}$; accurate numerical computations for tip fields and electron trajectories can be found in Scheinfein et al. [20].) In practice then, reducing the tip-to-sample distance z_1 increases the magnification (as $1/z_1$), reduces the defocus (see below), and reduces the voltage needed for the same tip field and emission current.

Figure 2 shows a low magnification image of a hole-containing carbon film obtained at about 400 V, using an instrument in which the tip motion along the axis is controlled by an inchworm.

2.2 Electron Ranges in Matter: Image Formation

Before discussing the image (hologram) formation process it is important to determine what fraction of elastically scattered 100-V electrons traverse a thin film of, say, carbon. This, together with the source size, will determine the nature of the scattering theory required, which might be based in a transport equation (if multiple inelastic processes dominate) or perhaps on transmission low energy electron diffraction (LEED) theory if the film thickness is much less than the inelastic mean free path (IMP). Since the beam energy is less than that of the inner shell atomic ionization energies, surface plasmons, phonon excitation, valence electron and plasmon excitation are the main energy-loss mechanisms. Extensive calculations of the inelastic mean free path λ_E have been made in the field of photoemission spectroscopy, and in support of LEED calculations [21]. This is defined such that the intensity of the elastic portion of the electron wavefield decays on entering a solid as

$$I(z) = I_0 \exp(-z/\lambda_E) = I_0 \exp(-2\sigma V_0' z) = I_0 \exp(-zN\sigma_t) \quad (1)$$

where $\sigma = \pi/\lambda U_0$ (λ is the electron wavelength), V_0' is the imaginary part of the mean inner potential (in volts), N is the density of atoms, and σ_t is the total inelastic cross-section. Measurements and calculations covering the range below 100 V (in which the theory becomes most difficult) can be found in the literature [22]. The results generally show a minimum of a few Angstrom for λ_E at about 80 eV for carbon, below which λ_E rises steeply with the onset of ballistic transport. Thus λ_E may be 100 nm or more near the Fermi energy for a pure metal at room temperature. An empirical formula has been proposed [23] of the form

$$\lambda_E = AE^{-2} + BE^{\frac{1}{2}} \quad (2)$$

where E is the electron energy in eV, λ_E is given in Angstrom, and A and B are fitting parameters. The values $A = 1430$ and $B = 0.54$ for elements, $A = 6410$ and $B = 0.96$ for inorganic compounds, and $A = 310$ and $B = 0.87$ for organic compounds are given. The traditional theoretical approach has been based on a modified Bethe theory of inelastic electron scattering. Another recent approach derives values of λ_E from measured optical dielectric functions $\epsilon(\omega)$ [24] using a model (such as the Lindhard dielectric function) for the dependence of scattering vector $\boldsymbol{q}$ on energy loss $\hbar\omega$. At low energies, exchange and correlation effects become important, and multiple inelastic scattering must be considered.

Figure 3 shows experimental measurements of the attenuation length L for low-energy electrons in the 7–1000 eV range [25]. These measurements were obtained using free-standing thin carbon films about 4 nm thick in transmission experiments, using an electron spectrometer to isolate the transmitted elastic scattering. L may be identified with λ_E under the approximation that the detector includes all scattering angles. Thus they are uniquely relevant to PPM imaging of organic films. Any undetected pinhole lying under the beam (of width about 60 μm) would increase the contribution of elastic scattering and so make this estimate of λ_E (about 0.6 nm at 100 V) too large. We see that thin films of this material only become reasonably transparent to electrons at energies below about 10 V. At energies around 100 V (where much point-projection work is done), the IMP is close to its minimum value. Since experience from TEM shows that it is extremely difficult to prepare extended areas of thin films less than 0.6 nm thick, we shall

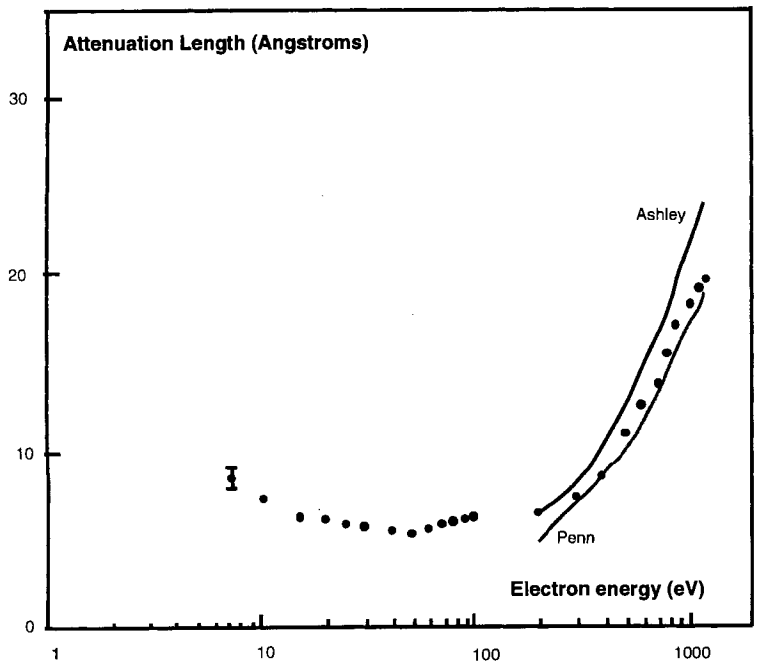


Figure 3. Measurements of the attenuation length L for low energy electrons in the 7–1000 eV range [23]. Dashed and continuous lines show theoretical estimates.

assume as an approximation that negligible transmission occurs, and give here the theory of point-projection image formation for masks, a mask being defined as a two-dimensional 'transmission' object whose transmission function is equal to zero (in opaque regions) or unity (in transparent regions). Such a theory may be useful for the important problem of determining the shape of molecules, if methods can be found for placing them, unsupported, across holes in carbon films. Clear experimental evidence for significant transmission of elastically scattered electrons at around 100 V is presently lacking. However, should such evidence be obtained, the theory of transmission LEED [26, 27] may be useful for edges of thin crystals, in which case Fourier imaging effects might also be observed. The theory of Fourier imaging in transmission at low voltage with multiple scattering is given by Spence and Qiau [28]. Experiments using an energy loss spectrometer attached to a PPM will clarify this point.

Figure 3 is consistent with the recent work of Shedd [29], who found that good agreement between simulated and experimental PPM images of carbon films with holes could be obtained by treating them as opaque masks containing holes, ignoring completely any partial transmission at edges. The two-dimensional patterns of fringes seen previously [14] were thereby found to be Young's fringes due to the interference between waves passing through three different pinholes, rather than atomic-resolution transmission lattice images. The two effects can easily be distinguished if the magnification (and hence the defocus) is accurately known. A more refined treatment would consider the contribution to the image of the large fraction of electrons which lose energy in traversing the sample, and the declining efficiency of channel plate detectors with electron energy below 100 V. An energy-filtering mesh has been fitted to a PPM instrument [14].

As discussed further in Sec. 2.7 of this Chapter, our experimental results suggest that unstained, uncoated purple membrane (thickness 5 nm) is opaque at 100 V, but reasonably transparent at 1 kV, and that significant transmission occurs through samples of the lipid monolayer C16 (thickness 2.6 nm) at 100 V. We are not able to determine the fractions of elastic and inelastic scattering in the transmitted beam.

We now consider the image-formation process in PPM. First, we demonstrate that a shadow image of a mask (or a weak phase object) is equivalent to a conventional out-of-focus image, and so discuss focus restoration schemes for masks. Consider a mask illuminated by a spherical wave originating from a source at distance z_1 from the mask. The transmission function for the sample is $q(x)$, but, for masks, we cannot assume the weak phase object approximation. Choose Cartesian coordinates with z along the beam path. Let the spherical wave incident on the sample be represented in the parabolic approximation by $t_{z_1}(x) = \exp(-i\pi x^2/z_1\lambda)$ (with Fourier transform $T_{z_1}(u) = C\exp(i\pi z_1\lambda u^2)$, where C is a complex constant) and let the electron wavefunction across the downstream side of the slab be $\Psi_i(x)$. Then, with $u = \Theta/\lambda$, where Θ is the scattering angle

$$\psi_i(x) = q(x)t_{z_1}(x) \tag{3}$$

the detected wavefunction in the far field

will be the Fourier transform

$$\psi_h(u) = \mathrm{FT}(\psi_i(x)) = Q(u)^* T_{z_1}(u) \qquad (4)$$

where the asterisk denotes convolution. Evaluating the convolution in Eq. (4), and ignoring unimportant phase factors gives

$$\psi_h(u) = \int [Q(U) \exp(i\pi z_1 \lambda U^2)] \times \exp(-2\pi i z_1 \lambda u U)\, dU \qquad (5)$$

Now, consider

$$z_1 \lambda u U = z_1 \lambda (\Theta/\lambda) U = z_1 \lambda (X/z_2 \lambda) U = (X/M) U,$$

where X is the spatial coordinate on the detector and $M = z_2/z_1$ is the magnification of the shadow image. Hence

$$\psi_h(X = z_2 \lambda u) = q(X/M)^* t_{z_1}(X/M) \qquad (6)$$

This important result establishes that, for masks (or for any thin sample for which a transmission function can be defined), the shadow image consists of an 'ideal' image which is out of focus by the tip-to-sample distance z_1. Equation (6) is identical to the expression for an out-of-focus high-resolution transmission electron microscope (HREM) image in the absence of lens aberrations. (Note, however, that for HREM plane-wave illumination is used.) However, the point-projection image has been magnified by $M = z_2/z_1$ without the use of lenses. We have compared simulated images based on Eqs. (5) and (6) and found them to be identical. As may readily be confirmed using an optical laser and a slide transparency, increasing z_2 increases the overall magnification of a shadow image, but not the focus defect, which is fixed by z_1. No assumption of periodicity has been made for the sample. An in-focus image can only be obtained using Eq. (6) if $z_1 = 0$, in which case M is infinite and, if a physical emitter is used, field emission then becomes impossible. (In addition, if $z_1 = 0$, zero contrast is predicted for the image of an in-focus transmission phase object.) The magnification and the focus setting of the PPM are not independent – for given z_2 both are fixed by z_1. For a crystalline sample for which the transmission diffraction orders overlap, the above analysis does not apply in the presence of multiple scattering, since then a transmission function cannot be used, as discussed elsewhere [30].

It follows from Eq. (6) that existing algorithms for focus correction (under plane-wave illumination) can be used to reconstruct low-voltage in-line holograms (formed with spherical-wave illumination). For masks, Fresnel edge fringes and Young's fringes between nearby holes are the main contrast features, and the aim of reconstruction is to remove these fringes, thus revealing the shape of the holes. The problem of focus correction (or hologram reconstruction) in the in-line geometry has been studied extensively, both for optical and electron holography. The first solution to the unavoidable twin-image problem which arises in this geometry may be that of Thompson and co-workers [31], who developed the optical method of in-line Fraunhofer holography for small particles. Solutions to the twin-image problem based on recording images at different defoci or lateral tip positions are described by Lin and Cowley [12]. Shadow images of masks have recently become important in the field of semiconductor lithography, where the inverse problem arises. For finer linewidths and sharper edges one requires the

edge transmission function whose defocused shadow image is most abrupt.

2.3 Holographic Reconstruction Algorithms

Given the PPM image (necessarily out of focus by z_1) of a mask-like object, the question arises as to whether the true shape of the mask can be recovered. For a mask we define the transmission function

$$\begin{aligned} q(x) &= 1 - p(x) \\ &= 1 \qquad \text{Within a hole} \\ &= 0 \qquad \text{Within opaque regions} \end{aligned}$$

Then, assuming the equivalence between point-projection shadow images and conventional images (Eq. 6), the wavefunction on a plane distance z_1 downstream from the sample is, for a magnification of unity [32],

$$\psi_h(x) = q(x)^* t_{z_1}(x) \tag{7}$$

The recorded hologram intensity is

$$\begin{aligned} I(x) &= \psi_h(x)\psi_h^*(x) \\ &= 1 - p(x)^* t_{z_1}(x) - p^*(x)^* t_{z_1}^*(x) \\ &\quad + |p(x)^* t_{z_1}(x)|^2 \end{aligned} \tag{8}$$

To reconstruct the object from the hologram, we first Fourier transform, then multiply by $T_{z_1}^*(u)$, and finally perform an inverse transform:

$$\begin{aligned} &\mathrm{FT}^{-1}\{T_{z_1}^* \cdot \mathrm{FT}[I(x)]\} \\ &\quad = 1 - p(x) - p^*(x)^* t_{2z_1}^*(x) \\ &\qquad + |p(x)^* t_{z_1}(x)|^2 * t_{z_1}^*(x) \\ &\quad = 1 - p'(x) \end{aligned} \tag{9}$$

The first two terms give a perfect reconstruction of the original transmission function $q(x)$, the real image. The third term gives the complement of the object, which is out of focus by twice the tip-to-sample distance z_1. If the object (of width d) is small, and z_1 large enough, this virtual twin image will produce a broad and slowly varying background which allows the real image to be isolated. [The requirement for this is $z_1 \gg d^2/\lambda$ (Fraunhofer holography).] The fourth term is a second-order term (actually the reconstruction of the hologram of the complementary object), which is small for weak phase objects, but not necessarily small for masks. Physically, this term might be expected to be troublesome in regions well inside the shadow edge, where the 'reference wave' is weak. The ratio of the wave scattered from the edge to the direct wave controls this term and, for a finite object, this ratio depends on the Fresnel number $N(d) = \pi d^2/(\lambda z_1)$. Thus reconstruction for masks introduces greater errors than for weakly scattering objects; however, for the purpose of identifying simple discontinuities at edges, we shall see that the Fraunhofer condition can be relaxed. We have investigated these issues using computational trials on simulated and experimental data [33].

In Fig. 4 we illustrate the use of the above algorithm for simulated data. The simulations require periodic boundary conditions (with supercell period L). Interference from neighboring cells can be avoided if the Fresnel number $N(L) \gg 1$. The reconstruction of two opaque square objects of width 1 and 3 nm (Fresnel numbers 0.43 and 3.86) is shown using 100-V electrons. The hologram was simulated using Eq. (8) and reconstructed using

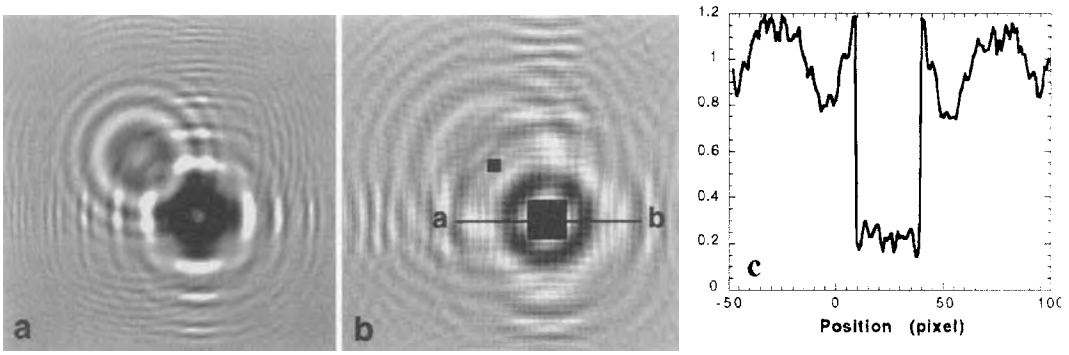


Figure 4. Simulation of data using Eq. (9): supercell period L, Fresnel number of $N(L) \gg 1$, 100 V electrons. Reconstruction [using Eq. (10)] of two opaque square objects of (a) width 1 nm; Fresnel numbers 0.43. (b) 3 nm; Fresnel number 3.86. Note how smaller square with smaller N is clearer. (c) The edge of the objects can be readily located from the profile.

Eq. (9). The edge of the objects can be readily located from the profile, despite the disturbance from the background.

These results suggest that any sharp discontinuity (such as an edge) in a finite object will produce a corresponding sharp discontinuity in the reconstructed image at Gaussian focus if N is not too large. This would allow the shape of objects to be determined with high resolution. We now apply this method to experimental data.

Before attempting reconstruction, the magnification $M = z_2/z_1$ of the images must be known. Five methods have been used:

(1) Use of a through-magnification series and an object (such as a grid bar) of known size. This is very inaccurate since errors between successive images compound.
(2) Use of parallax-shift of images with lateral tip movement [15].
(3) Analysis of Fresnel edge fringes [34].
(4) Use of Young's fringes observed between two pinholes of spacing D, using $d = z_2\lambda/D$, where d is the fringe spacing and z_2 is the sample-to-screen distance.
(5) Trial-and-error values of z_1 may be used in the reconstruction scheme until a sharp image is obtained, as in Fig. 4b.

We have used methods (3)–(5).

Figure 5 shows two experimental point-projection holograms of a hole-containing carbon film obtained at 90 V. Figure 5b

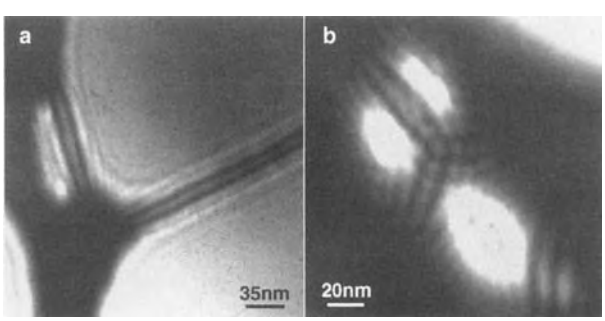


Figure 5. Experimental point-projection holograms of a hole-containing carbon film obtained at 90 V. (a) In-focus image. (b) Crossed sets of Young's interference fringes between three pinholes.

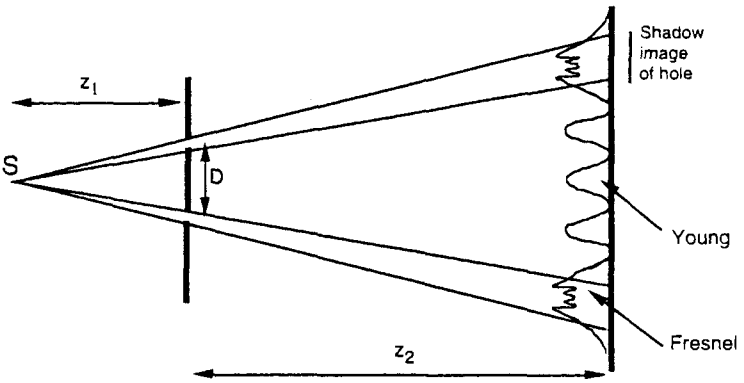


Figure 6. In the near-field of a mask containing holes, both a shadow image of the holes and interference fringes are formed.

shows the crossed sets of Young's interference fringes between three pinholes which might be confused with an atomic-resolution lattice image. The fringes may be simply interpreted using Eq. (6), expressed in two dimensions. Then we see that the fringes are equivalent to near-field Young's fringes, as would be obtained using plane-wave illumination. In the near-field of a mask containing holes, both a shadow image of the holes and interference fringes between different holes are observed [29], as indicated in Fig. 6.

The holograms shown in Fig. 5 were also reconstructed using Eq. (9) and a program which allows for the change of object pixel size with variation of defocus z_1, which changes the magnification. An approximate value of z_1 was assumed and the corresponding magnification $M = 14\,\text{cm}/z_1$ used. Since the computations are fairly rapid, the effect of varying the trial value of z_1 simulates changing experimental focus near Gaussian focus, and by comparing reconstructed amplitude and phase images, the in-focus image may readily be identified by eye, despite the background from the twin image and non-linear term. The focus correction needed to obtain the in-focus image is the experimental tip-to-sample distance.

The shape of the fibers retrieved from the in-focus image reconstructed from Fig. 5a is shown in Fig. 7a, while Fig. 7b shows

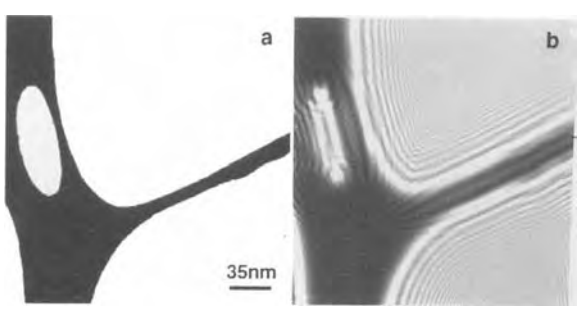


Figure 7. (a) Shape of the fibers retrieved from the in-focus image reconstructed from Fig. 5a; (b) forward simulation of the hologram based on (a).

a forward simulation of the hologram based on Fig. 7a, in excellent agreement with the experimental hologram (Fig. 5a) if a defocus of 1850 nm is used. An integration over the finite source size would improve the fit by washing out the higher order Fresnel fringes. Reconstructions obtained from Fig. 5b eliminate entirely the Young's fringes at certain focus settings [33].

The sharpness of the edges in the reconstruction depends on the Fresnel number $N = \pi d^2/\lambda z_1$, which describes the effect of defocus z_1, wavelength λ, and object diameter d. For the edge of an 'infinite' halfplane (N infinite) the reconstruction is noticeably poorer than for a small object. We find that for a simple mask object the Fraunhofer condition $N \ll 1$ for elimination of the twin image is an excessively stringent requirement. In addition, any phase shift on scattering at edges, or partial transmission, will also affect the reconstruction.

In summary, we find that any sharp discontinuity (such as an edge) in a finite object will produce a corresponding sharp discontinuity in the reconstructed image at Gaussian focus if N is not too large. As Fig. 4 shows, this focus setting (the tip-to-sample distance) can easily be identified from a focal series of reconstructed images. This discontinuity, identifiable against the background of the twin image and the non-linear terms, easily allows the outline shape of small opaque objects to be identified. Resolution in the reconstructed image will still be governed by the electron source size (which determines the highest order Fresnel fringe, and hence the size of the hologram, since the Fringes get progressively finer according to Eq. (13)).

2.4 Nanotips, Tip Aberrations, Coherence, Brightness, Resolution Limits, and Stray Fields

The observation of Fresnel fringes (i.e., the ability to form a hologram) in the point-projection mode requires either demagnification of a conventional electron source (as in the original experiments of Boersch and in STEM work), or a physical source of subnanometer dimensions. This may be obtained either by using naturally occurring asperities (as in STM), or by preparing an atomic structure of desired shape. An extensive literature exists in the field of FIM on tip preparation procedures, based on oxygen etching, field evaporation, heating or sputtering [35]. More recently, using FIM imaging to assist in characterization, single-atom and few-atom 'nanotip' field emission tips have been prepared [15], for which claims of very high brightness and other unique properties have been made, including departures from the Fowler–Nordheim law, 'focusing' effects and unusual emission energy spectra [36]. Theoretical treatments for diffraction of a wavepacket at a laterally confined tunneling barrier have also been given (for a review of all this work see Vu Thien Binh et al. [37]). We now discuss the nanotip properties most relevant to PPM, their aberrations and brightness. Figure 8 shows an idealized tip and the curved trajectory of electrons accelerated from the tip. Since most of the potential is dropped within a few hundred nanometers of the tip, the electrons reach their final kinetic energy rapidly. The rearward

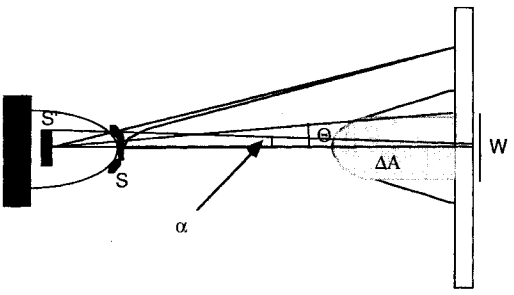


Figure 8. Idealized tip used in PPM and curved trajectory of electrons accelerated from the tip.

asymptotic extension of these rays defines the virtual source inside the tip, whose size d depends on the failure of rays leaving at different angles to meet at a point in the Gaussian image plane inside the tip (defined by paraxial rays). Thus the tip acts as a lens of nanometer dimensions, whose aberration coefficients can therefore be expected to be also of nanometer dimensions. Aberration coefficients for such a nanotip have been computed by modern ray-tracing techniques, together with numerical solutions of Laplace's equation [20]. For a tip with 1 nm radius sitting on a boss with 100 nm radius, we find $C_s = 0.177\,\text{nm}$ and $C_c = 0.142\,\text{nm}$ at 100 V, with a tip field strength at the surface of the tip of $5\,\text{V nm}^{-1}$. We shall see that the effects of these aberrations on resolution may not be negligible, because the virtual source size depends also on emission angle and wavelength. The emission angle θ for this tip was found to be 7.4° (halfwidth at e^{-1}).

From Fig. 8 the coherence width at the detector is seen to be approximately $X_c = \lambda/\alpha$. It is also useful to define a coherence angle $\beta = \lambda/d$, which is equal to the angular range over which emission from the tip is coherent – this is the angle subtended by X_c at the tip. The virtual source diameter d for the 1 nm radius nanotip discussed above is found to be 0.43 nm (i.e., demagnified). For such a source the coherence width on a sample at $z_1 = 100\,\text{nm}$ is 54 nm, while that on a detector at $z_2 = 10\,\text{cm}$ is almost 6 cm ($\lambda = 0.126\,\text{nm}$, 100 V). The coherence angle is greater than the emission angle. The coherence width is approximately equal to the width of all the Fresnel edge fringes which can be seen, and hence to the size of the hologram. Because of their high brightness, nanotips may therefore be useful for electron interferometry experiments in which the beam must pass around macroscopic objects smaller than X_c. Measurements of source size have recently been made for a nanotip using a biprism [38]. The longitudinal coherence length $L_c = \lambda(E/2\,\Delta E) = 63\,\text{nm}$ for a nanotip with $\Delta E = 0.1\,\text{eV}$ at 100 V is considerably less than for high-voltage field emission guns.

The brightness $B = j/\Delta\Omega$ (see Fig. 8) of a nanotip has been measured by matching ray-tracing computations and Fowler–Nordheim curves to experimental nanotip intensity distributions [39]. By comparing field ion microscopy and field emission electron images with the ray-tracing results, the effective source size could be obtained. Tungsten (111) single-crystal tips were fabricated using Ne sputtering and field evaporation. The average brightness of single-atom-terminated nanotips was found to be $3.3 \times 10^8\,\text{Å cm}^{-2}\,\text{sr}^{-1}$ at 470 V, or $7.7 \times 10^{10}\,\text{Å cm}^{-2}\,\text{sr}^{-1}$ when extrapolated to 100 kV. This produces a beam with greater particle flux per unit energy range than obtainable using current synchrotrons. Although this value is almost two orders of magnitude brighter than the values quoted for STEM

cold-field emission guns, it should be borne in mind that the STEM values are based on measurements of the focused probe intensity, which is affected by objective lens aberrations. When corrections are made for this effect, the nanotip brightness is found to be greater than cold-field emission by a factor of about 10 at 100 kV. The most valuable aspect of the nanotip for STEM instruments may be the increased emission stability obtained at low extraction voltages, the reduced emission energy spread ΔE, and the increased tolerance to poorer vacuum conditions. It has been speculated [40] that this is due to the reduced energy of backstreaming ions, which at higher extraction voltages may modify the workfunction of the tip.

Source brightness can be measured more directly using the concept of degeneracy $\delta = B/B_0$, where B_0 is the upper quantum limit on brightness corresponding to two Fermions per cell in phase space (if $\delta = 1$). The degeneracy can also be expressed as

$$\delta = (j/e)A_c T_c \tag{10}$$

This is the mean number of particles per coherence time $T_c = h/\Delta E$ traversing a coherence area $A_c = X_c^2$ (normal to the beam). In practice, this means that δ (and hence the brightness) can be measured if the beam energy spread and the current density within a patch of Fresnel fringes are known, since the width of these fringes is given approximately by X_c. δ is unity if $j = e/(A_c T_c)$, so that one electron crosses the coherence area per coherence time. Using the expression for the constant B_0, it follows [34] that the experimental brightness can be determined from the wavelength, the current density j and the coherence area at the image plane using

$$B = 2jA_c/\lambda^2 \tag{11}$$

The effect of source size on resolution may be understood by considering the effect of displacing an ideal emission point transversely by b. This translates the entire image by Mb, where $M = z_2/z_1$. This imposes a resolution limit on the images, since the complete image intensity must be integrated over the effective source size d, which cannot be smaller than the electron wavelength. This integration takes into account all partial spatial coherence effects. For masks, the resolution of the PPM is most conveniently defined operationally as the width of the finest Fresnel fringe (in the object space), since this is readily shown to be equal to the size of the electron source (including the effects of all instabilities), as we now show. It was first shown by Sommerfeld that a good approximation to the Fresnel integral may be obtained by summing over just two optical paths (rather than infinitely many), the direct ray from source to detector, and that which passes via the edge of the mask. This gives the correct position of maxima and minima, but not the intensity distribution, if path differences $\Delta_0 = 3\lambda/8$ for the edge wave scattered outside the shadow edge and $\lambda/8$ for the edge wave scattered inside are used. Then the lateral position X_n of the nth Fresnel fringe on the screen is

$$X_n = M\sqrt{2z_1(n\lambda + \Delta_0)} \tag{12}$$

The width of the nth fringe on the screen is

$$\delta X_n = \frac{M^2 z_1 \lambda}{X_n} \tag{13}$$

Setting $X_c = X_n/M$ at the sample we find the width of the highest order fringe at the

sample is approximately equal to the source size $d = 2z_1\lambda/X_c$. In practice, however, the contrast of the fringes decreases with order, making measurements of source size and coherence by this method approximate. (For a full analysis of Fresnel fringes in PPM see Spence et al. [34]. The image magnification, effective source size, transverse coherence width, instrumental resolution, and source brightness are all obtained from an analysis of the fringe spacings and intensity.)

The resolution limit in PPM may be thought of in other ways. By limiting the size of the hologram, the coherence width X_c imposes a diffraction limit on the reconstruction, since the hologram acts rather like a lens (or zone-plate) during reconstruction. Thus the coherence angle β plays a similar role to the angular limit set by the aperture of the objective lens in HREM (as suggested by θ in Fig. 1) – a larger angle is needed for higher resolution d_r. The diameter of the zone plate is approximately equal to the transverse coherence width MX_c at the detector screen. Thus $d_r = 1.6\lambda z_2/MX_c = 0.8\lambda/\beta$, in agreement with the idea that the coherence angle from the source acts as a diffraction-limiting aperture. Higher resolution is obtained by increasing the emission coherence angle $\beta = \lambda/d$, and this can be achieved by reducing the source size d. A useful result is obtained by noting that, since $X_c = 2z_1\lambda/d = \lambda/2\alpha$ (where α is the angle subtended by the source at the sample) and, since the resolution is $d_r = 1.6\lambda z_1/X_c$, then $d_r = 3.2z_1\alpha$. Thus resolution in the reconstruction is improved by using small tip-to-sample distances and the smallest sources. For a thin crystal, we may again think of α as limiting the 'numerical aperture' of the nanotip lens. Then the finest fringes contributing to a Fourier image will be those with Bragg angle $2\Theta_B = a$ for plane spacing d_{hkl}. Using $l/d_{hkl} = 2\Theta_B/\lambda$, we again see that the resolution is approximately equal to the source size $d = d_{hkl}$.

The effects on resolution of the spherical aberration of the tip must be considered. Taking the resolution to be very crudely equal to the virtual source diameter Δ_{C_s}, we have, from the definition of spherical aberration,

$$\Delta_{C_s} = 2r_i = 2M(C_{s3}\sin^3\theta + C_{s5}\sin^3\theta) \qquad (14)$$

where θ is measured in the object space (the emitting surface) and, from calculations, $M = 0.531$ at 100 V. For emission at $\theta = 45°$, $\Delta_{C_s} = 0.84$ Å for the 1 nm radius nanotip. It is instructive to compare this with the performance of a magnetic lens for high-energy electrons. Here, at 100 kV ($\lambda = 0.037$) one might typically have $C_s = 0.5$ mm, but an angular aperture set at about $\theta = 10$ mRad. Then $\Delta_{C_s} = 2.5$ Å because of the much smaller angles but larger aberrations. The idea that the aberrated virtual source size is equal to the resolution, however, is not strictly correct, since no simple expression for the resolution-limiting effects of spherical aberration on shadow images has been given, even for weakly scattering objects [41].

The effects on resolution of mechanical vibration of the tip and an enlarged emission area may be included by summing the image intensity distribution given by Eq. (5) over a range of source points using a suitable source intensity distribution function. This incorporates all partial spatial coherence effects. It is instructive to compare the effects of tip motion in several

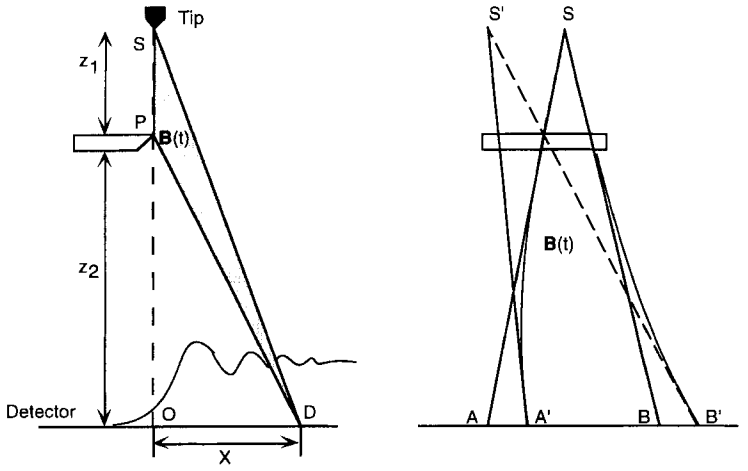


Figure 9. A transverse time dependent magnetic field will introduce a time dependent phase shift between the optical paths SPD and SD (left); construction for estimating the degrading effect of stray magnetic fields on the resolution of a general PPM image (right).

point projection microscopies. For both electron field emission microscopy and FIM, the image is a projection of the surface of the tip from a virtual source inside the tip which cannot move relative to the surface of the tip. The effects of vibration in these instruments are therefore not severe, since they consist of whole-body translations of the source and 'sample' together, which are not magnified. A 1-nm movement of the tip appears as a 1-nm displacement on the screen. For the PPM and STEM in-line holography, any independent vibration at the source appears magnified at the detector. It is perhaps this relative immunity of the field ion microscope to vibration effects which accounts for the fact that atoms were first seen with these instruments. For conventional microscopy (TEM) it is the relative motion of the sample and the detector (rather than the source) which is crucial, since plane-wave illumination is used.

The effects of a spread of energies ΔE in the beam has been discussed previously [34]. For Fresnel fringes, it was found that energy broadening of the source imposes a chromatic resolution limit such that the highest order fringe which may be observed has order $n_{\max} = 2E/\Delta E$. The width of this fringe (referred to the object space) is

$$\Delta C_r = \sqrt{z_1 \lambda/(2n_{\max})} \tag{15}$$

which, by substituting for $n_{\max}$, gives the resolution limit due solely to the energy spread of the beam. For $z_1 = 1000\,\text{nm}$ and $\Delta E = 0.1\,\text{eV}$ we obtain $\Delta C_r = 0.17\,\text{nm}$ at $E = 100\,\text{eV}$.

As shown in Fig. 9a, a transverse time-dependent magnetic field will introduce a time-dependent phase shift between the optical paths SPD and SD. This phase shift is

$$\theta(t) = \frac{e}{h}\int \boldsymbol{B}(t) \cdot d \tag{16}$$

where the integral is carried out over the shaded area in the figure. The use of just two optical paths rather than the Fresnel integral can be shown to give the correct positions of the maxima and minima in the Fresnel fringes. This phase shift may be included in the Fresnel fringe analysis, and a time average of the intensity taken in order to compute the effects of stray fields.

The result will depend sensitively on the order of the fringe. A rough estimate may be made by setting the maximum allowable phase shift equal to $\pi/4$ at the position of the nth Fresnel fringes and solving for B. This fringe occurs at $X_n = M[2z_1(n\lambda + \Delta_0)]^{\frac{1}{2}}$ on the detector, with $\Delta_0 = \frac{3}{8}\lambda$ the approximate phase shift on scattering at the edge. This procedure gives

$$\langle B \rangle_{\max} = \frac{\pi}{4} \frac{h}{e} \frac{2}{z_1 X_n} \tag{17}$$

The fringe width δX_n at the detector is related to X_n by $\delta X_n = M^2 z_1 \lambda / X_n$. The smallest (highest order) fringe is equal to the incoherent virtual source size. If we take this to be, say 1 nm (referred to the specimen plane), then about six fringes will be seen. The maximum tolerable time-dependent field which allows observation of the sixth fringe is then $\langle B \rangle_{\max} = 6 \times 10^{-7}$ T (a few milligauss) at 100 V with $z_1 = 100$ nm and $z_2 = 14$ cm. Only the component of the field normal to the shaded area is important. The constant Earth's field (about 0.6 T) can be ignored.

Figure 9 also suggests a construction for estimating the degrading effect of stray magnetic fields on the resolution of a general PPM image. If a homogeneous field $\boldsymbol{B}(t)$ fills the space between sample and detector in the direction shown (worst case), the electron's trajectory will be circular, and the rearward asymptotic extension of these deflected rays at the detector will define a displaced source point S′. The distance S–S′ gives a measure of the source enlargement, and hence resolution loss, due to a time dependent field. We note that S–S′ also contains a z-component, and so there is a defocusing effect.

For steady fields, the angular deflection of a non-relativistic electron of energy eU_0 acted on over path length L by a transverse electric field F is

$$\theta = \frac{FL}{2U_0} \tag{18}$$

while for a magnetic field with component B_t transverse to the beam the deflection is

$$\theta = \left(\frac{e}{2m}\right)^{\frac{1}{2}} \left(\frac{B_t L}{U_0^{\frac{1}{2}}}\right) \tag{19}$$

In summary, the width of the finest Fresnel fringe from a carbon foil edge is the most useful practical measure of resolution for the PPM. Energy-broadening and source-size effects may limit resolution to approximately the electron wavelength. They do not impose an absolute limit, however, and, as in near-field and other forms of microscopy, resolution will ultimately be limited by the signal/noise ratio. The non-linear terms in holographic reconstruction methods also limit resolution. In the presence of multiple scattering, resolution cannot easily be defined. Resolution increases with emission coherence angle, which acts as a diffraction-limiting aperture, as for a lens. This angle $\beta = \lambda/d$ (where d, the effective source size, is a measure of the resolution) can be measured using the methods outlined above [34] or, better, by using an electron biprism [38].

2.5 Instrumentation

Several designs for low-voltage point-projection instruments have appeared in the literature. All consist of a tip motion, a sample holder and a detector, for which all

recent instruments have used a single-stage channel-plate followed by a fluorescent screen. These items are supported within an ultra-high vacuum chamber which must be bakeable (at least initially) and supported on a vibration isolation system. A leak valve for sputtering and vacuum gauges may be required. Pumps must pump the noble gas used for sputtering. Fine and coarse motion of the tip relative to the sample along x, y, and z directions is essential to bring a region of interest onto the optic axis, defined by the beam, and to control the magnification and focus. Tip motions have been based on the original piezo tripod scanner used in STM [14] or on a piezo tube scanner for lateral motion and an inchworm for coarse z motion [18]. Tip cooling to 4.2 K has been provided [19]. Tip heating (for cleaning) may be provided by resistance heating (in which case the tip support follows the design of the tip assembly in field-emission STEM instruments), or by indirect electron-beam heating, which achieves lower temperatures but allows a higher mechanical resonant frequency for the tip, and does not require thick high-current leads to the tip assembly. If field-ion imaging of the tip is planned, electrical insulation for voltages of at least 10 kV must be provided to the tip. Since resolution is degraded by tip vibration, resistance heaters must be short and rigid, and all the principles of good STM design apply [42]. For a machine supported on a spring system with natural resonance frequency f_{sp}, the attenuation K of vibration amplitude reaching a tip with natural resonance frequency f_{tip} is approximately

$$K = \left(\frac{f_{sp}}{f_{tip}}\right)^2 = \frac{a_{tip}}{a_{ext}} \qquad (20)$$

if $f_{sp} \ll f_{tip}$. Here a_{tip} is the vibration amplitude of the tip, while a_{ext} is the amplitude supplied to the frame of the support. The frequencies are proportional to the square root of the relevant spring constant divided by a mass. One makes K as small as possible by making f_{sp} as small as possible (e.g., by using an air-table or elastic suspension system with large mass and small spring constant), and f_{tip} as large as possible (by making the tip as short and rigid as possible). Typical values of $f_{tip} = 2$ kHz and $f_{sp} = 2$ Hz give $K = 10^{-6}$ or 120 dB attenuation. Tips with f_{tip} above 10 kHz are highly desirable. A short mechanical path between tip and sample is required to minimize relative displacements of the tip and sample – a translation of both together is not magnified.

As for STM, the most difficult design problem is the coarse lateral motion of the sample relative to the tip. A design has been described [18] which provides one-dimensional transverse motion of a carriage riding on three sapphire balls, based on an inertial stick-slip motor. This, however, requires careful leveling to avoid consistent downhill motion. The use of a molybdenum plate riding on sapphire balls has been reported to solve this problem. (Unlike STM stick-slip stages, magnets cannot easily be used to stabilize the motion against gravity.) Transverse inchworm motion has been used; however, jumps in the motion and the loss of a short, rigid mechanical path between tip and sample may limit resolution. The crucial requirement is to be able to control *both* z_1 and the coarse lateral stage motion interactively, in order to center a feature of interest on the optic axis. This is simple once one is within range of the fine lateral motion provided by a tune scanner

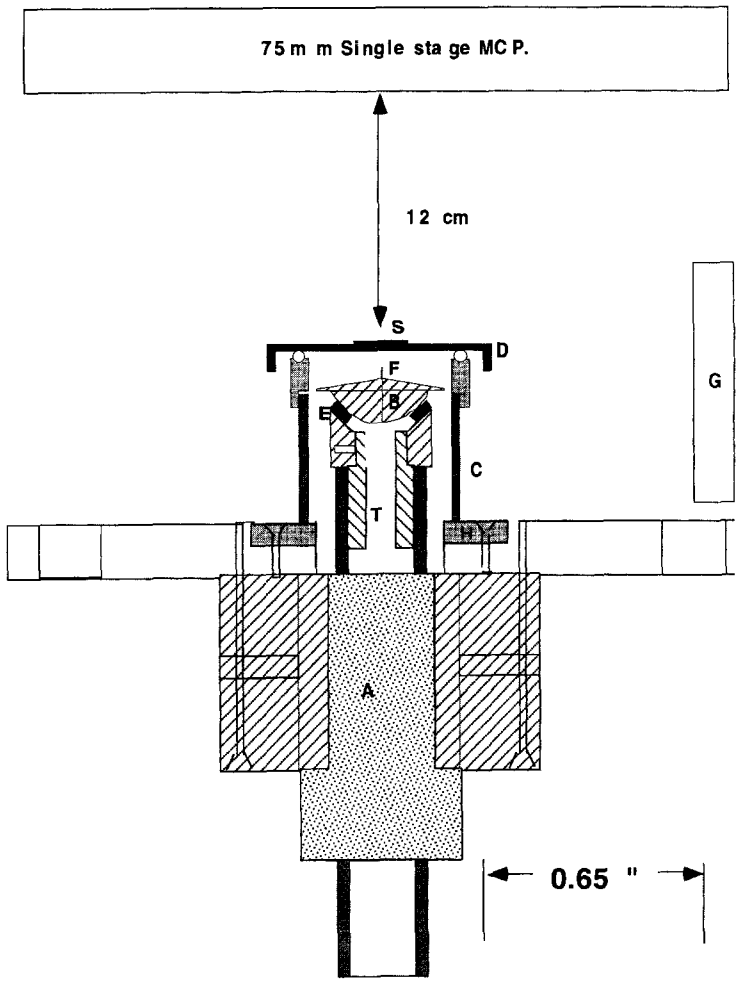


Figure 10. A recent vertical axis PPM which allows convenient tip and sample exchange, together with a stick-slip goniometer [43] for alignment of the beam.

supporting the tip. High-stability, low-noise electrical supplies (e.g., batteries) may be required for the scanner, since lateral sensitivities of $50\,\mathrm{nm\,V^{-1}}$ are common.

Figure 10 shows a recent vertical axis design [43] which allows convenient tip and sample exchange, together with a stick-slip goniometer [44] for alignment of the beam. This allows one to take advantage of natural asperities on a single-crystal tungsten tip, which, in our experience (and that of STM workers) can be extremely sharp in the sense of producing very fine Fresnel edge fringes. The goniometer is also used for experiments in the reflection mode. Since natural asperities are unlikely to be aligned with the [111] axis, the goniometer allows them to be rotated onto the axis. This use of chemically etched tips without heating, sputtering or field evaporation then allows organic samples, which cannot be baked, to be studied at 10^{-8} T. Lower vacuum

produces excessive ion background in the MCP. In Fig. 10, A is an inchworm and E is three shear-mode piezo stacks which rotate the hemisphere B about two orthogonal axes by the stick-slip mechanism, thus aligning the tip F eucentrically. The stage D (with samples S on a TEM grid) sits on three balls on top of a piezo outer tube C, whose four electrodes are driven by a ramp, giving lateral fine or (stick-slip) coarse motion of the sample. A pumped load-lock at G allows either the sample to be removed using a pick-up fork, or, subsequently, after extending the inchworm upward, the hemisphere containing the tip may be exchanged by a smaller fork. The previously baked UHV environment within is thus preserved. An internal μ-metal cylinder surrounds the instrument, which sits on a 60 l/s ion pump.

A simple technique for initial chemical etching of tungsten tips has been described which allows optical examination during etching [45]. Such a tip will field-emit at about 600 V or less initially, and this voltage may fall to 50 V or so as the tip is brought to within a few tens of nanometers of the conducting sample. One of the most important findings from nanotip research is the unexpected stability of field-emitters at low voltage, and their tolerance to rather poor vacuum conditions. It has been suggested [40] that this occurs because electrons striking the 'anode' (the sample) have insufficient energy to sputter ions efficiently, which are normally focused back onto the tip, causing local modifications in the workfunction. By thus avoiding the need for baking after each sample exchange, the study of organic films is made possible using a nanotip field emitter.

2.6 Relationship to Other Techniques

We have seen that the point-projection method has its origins in the very beginnings of electron microscopy, where much higher voltages were used. Many of the effects seen at higher voltage can also be expected in the PPM if a transmitted (or reflected) beam can be obtained, using any of the methods listed in Sec. 2.7 of this Chapter. The central disk of an out-of-focus coherent convergent-beam pattern may be interpreted in an in-line electron hologram if the sample is sufficiently thin (for a review see Spence and Zuo [46]). The theory of STEM holography and shadow imaging using 100-kV electrons is given elsewhere [11, 40], and this is closely related to the theory of lattice imaging in STEM, which depends on interference between coherent overlapping diffraction orders [47, 48]. The resulting fringes may be interpreted as a shadow image of the crystal lattice [46], and such images have been obtained at atomic resolution [49]. For crystalline samples, a family of real or virtual sources (probe images) is formed by Bragg diffraction, lying on the reciprocal lattice and forming a point diffraction pattern [46] (exactly as in the Tanaka wide-angle CBED method, which isolates one of these). If the illumination aperture is much larger than the Bragg angle, the patterns are known as 'ronchigrams' and provide a useful aberration figure of the probe-forming lens, as described in Sec. 2.2, Chapter 4 of this Volume. Point projection has also been used as an imaging method using ions [50] and X-rays [51]. Possibilities using matter waves in general (e.g., atom interferometry or neutrons) are

discussed by Clauser and Reinsch [52]. Since spin effects between different beam electrons are negligible, an exact analogy exists between the theory of in-line optical holography and electron PPM imaging of masks, so that, as we have seen, the methods of Fraunhofer optical holography [31] may be used to solve the twin-image problem for small objects. Other approaches to solving the twin-image problem have been proposed [12].

2.7 Future Prospects, Radiation Damage, and Point Reflection Microscopy

The present generation of instruments are capable of providing subnanometer-resolution images which reveal the external shape of mask-like objects. They may thus be useful for determining the shape of molecules drawn across holes in thin carbon films. In order to obtain images of internal atomic structures, one must use either higher or lower voltage (see Fig. 3).

Higher voltages become possible if the tip field is shielded by a cup-shaped surround, so that a higher voltage is needed for the same tip field. This shield (which is actually a lens) must not obstruct the sample. Lithographic techniques may allow such structures to be formed. Withdrawing the tip and using lenses after the sample to compensate for lost magnification allows higher voltage to be used (up to about 1 kV), but the image then has a very large focus defect. The aberrations of a second stage lens would be unimportant because of the small angles involved. Our experience has been that purple membrane (thickness about 5 nm) becomes reasonably transparent to electrons above about 1 kV. Using a blunt tip, lower-resolution transmission images have been obtained at this voltage.

Research in surface science indicates that greater penetration should be obtainable at much lower voltage (see Fig. 3). Preliminary PPM experiments down to 7 eV have been encouraging [17]. Wedges of silicon may be useful test samples, and the resonances seen in LEED and elsewhere should be observable. Image interpretation becomes extremely complex at very low voltage (since exchange and correlation effects become important) and the increasing electron wavelength may limit resolution, unless near-field conditions apply. Full band-structure computations may be needed to interpret the images.

An important question concerns the effects of radiation damage at low voltage. The Bethe stopping power law predicts a cross-section for inelastic scattering increasing as $1/v^2$ as beam energy decreases, but not all excitations result in damage, and this law fails below the ionization energy threshold. Rather little is known regarding damage in transmission samples at low energy; however, there are indications that it may decline in some systems [53]. Figure 11 shows a PPM image of a cut in a film of unstained, uncoated purple membrane (bacteriorhodopsin) showing no damage at 2-nm resolution after long exposure [54]. However, damage may have saturated in a surface layer, and samples of this thickness (5 nm) then still act as masks at 100 V. By contrast, for C16 lipid monolayers (thickness 2.6 nm) significant transmission was obtained at 100 V (perhaps mostly inelastic

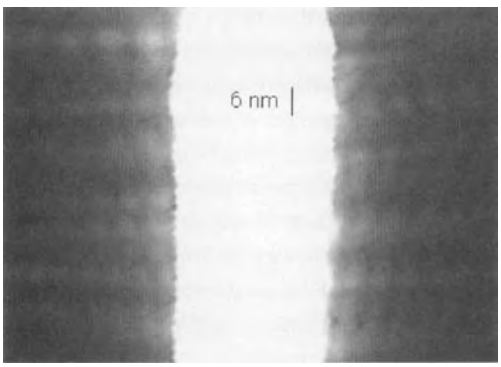


Figure 11. PPM image of a cut in a film of unstained, uncoated Purple Membrane (bacteriorhodopsin) showing no visible damage at the 2 nm resolution level after long exposure [53].

scattering) and damage also observed. The interpretation of the purple membrane image in Fig. 11 has been confirmed by image simulations based on Eq. (5). The important point is that the fringes running normal to the edges of the vertical cut in the film can be explained by roughness at the edges alone, and do not indicate significant transmission of electrons. Because this roughness reflects the periodicity of the crystal, the fringes also show the period of the crystal [55].

It has been suggested that inner-shell ionization may be the primary cause of damage at higher voltages in some materials [56]. Although less probable than valence excitation, these processes involve larger, more localized energy transfers. Then damage would be greatly reduced at beam energies below the carbon K ionization energy (284 eV). (In fact, competition with other loss processes and multiple scattering may raise this threshold appreciably [53].) Secondly, a recent study of the energy dependence of damage found a distinct threshold for aromatic (but not aliphatic) compounds, such that damage fell abruptly below about 1 kV [53]. Finally, it is commonly found [57] that radiolysis following valence excitation is suppressed in aromatic macromolecules with low G factors rich in π bonds (such as the nucleic acids), due to the delocalized nature of these resonant states, or by rapid screening of the hole on a time-scale faster than any nuclear response. A study of energy-loss spectra obtained from PPM images at various beam energies would contribute greatly to our understanding in this poorly understood and complex field. The elastic and inelastic cross-sections are comparable for electron beams at low energy, but inelastic scattering dominates for X-ray scattering. We therefore expect much less radiation damage per unit information using point projection electron methods than using comparable X-ray holography methods (see [58] for a detailed comparison).

More useful images might be obtained in the reflection geometry from bulk crystals with atomically smooth surfaces. The instrument shown in Fig. 10 was designed to investigate this possibility, using the tip goniometer to control the beam direction. Two methods have been proposed – one based on the Lloyd's mirror geometry [59] and the other using the specular reflected cone of radiation from a 'point' source [60] (i.e., RHEED with a spherical incident wave). The second method may be demonstrated using a diverging laser beam reflected from a mirror, on which is placed a small object. Both reflection high-energy electron diffraction (RHEED) computations (at a few hundred volts) and ray-tracing calculations [61] suggest that useful shadow images of surface steps can be obtained in this way if the experimental conditions can be controlled accurately.

In these low energy multiple scattering calculations, the results for reflection from an atomically smooth surface are modified to include the effect of a surface step using a simple phase shift. The experimental method relies on the fact that the electric field falls off rapidly near the tip, so that electron trajectories are relatively unaffected by the orientation of a reflecting surface a few micrometers away. Reflected rays appear to come from a virtual source below the mirror-like surface, whose distance from the surface defines the defocus. A diffusely scattering object on the surface scatters electrons weakly which interfere with the specular RHEED beam, forming a shadow image and hologram at infinity, as shown elsewhere [60, 61]. Multiple scattering computations show that the image is modulated by the convergent-beam RHEED rocking curve for the specular beam. Preliminary results are encouraging, and we note that similar shadow images can be seen in defocused reflection convergent beam patterns. Details are given elsewhere [61]. The first experimental low-voltage point reflection images were obtained recently, and are described elsewhere [62]. This instrument holds promise as a small, lensless, inexpensive alternative to the more complex low-energy electron microscope (LEEM).

In summary, the most promising areas of research in this field appear to be as follows.

(1) The development of imaging energy filters to exclude inelastic scattering. A filter of the imaging retarding Wien type may have advantages [63]. We note that the declining detective quantum efficiency of channel plates below about 100 V has a filtering effect.

(2) Development of the reflection mode [62]. If successful, this could be applied to problems of crystal growth in surface science and magnetic and ferroelectric domain imaging, as in the fields of (REM) and STM.

(3) Development of methods for preparing self-supporting molecule assemblies, perhaps drawn across holes in carbon films. The important problem of determining molecular shape for molecules which cannot be crystallized might be tackled in this way with existing instruments. In order to avoid the restriction to small objects ($N \ll 1$), we are exploring the use of recursive calculation between two defocii, and the imposition of the a priori constraint that $q = 1$ or 0 in image processing. The use of lower voltage (where IMF increases) may increase transmission, but reconstruction is needed to test this, since defocus produces delocalized images, and Young's fringes between different pinholes may otherwise give a false impression of transmission.

(4) Development of lithographically formed tip assemblies (mini-lenses) to enable operation at higher voltages, allowing greater penetration, but still below the inner shell ionization energies. For the study of lipid monolayers and Langmuir–Blodgett films, a compromise voltage might be found; however, it is to be expected that the effects of radiation damage will be highly materials dependent.

(5) For research in electron interferometry, electron anti-bunching, electron holography and the Aharanov–Bohm effect, the PPM provides an inexpensive experimental arrangement with excellent counting statistics due to the high source brightness. Chromatic coherence, however, is limited. Nanotips may also be

useful for electron spectroscopy in view of reports of very low energy spreads in the emission distribution [36].

Acknowledgements

I am most grateful to Drs W. Qian, J. M. Zuo, X. Zhang and K. Taylor for their assistance with this project which was supported by NSF award DMR 95-26100.

2.8 References

[1] H. Boersch, *Z. Techn. Physik.* **1939,** *12*, 346.
[2] H. Boersch, *Naturwissenschaften* **1940.** *28*, 709.
[3] J. Hillier, *Phys. Rev.* **1940,** *58*, 842.
[4] D. Gabor, *Proc. R. Soc., Ser. A* **1949,** *197*, 454.
[5] M. E. Haine, T. Mulvey, *J. Opt. Soc. Am.* **1952,** *42*, 763.
[6] T. Mulvey, Personal communication **1994**.
[7] G. A. Morton, E. G. Ramberg, *Phys. Rev.* **1939,** *56*, 705.
[8] A. J. Melmed, J. Smit, *J. Phys. E* **1979,** *12*, 355.
[9] J. M. Cowley, A. F. Moodie, *Proc. Phys. Soc. B* **1957,** *70*, 486.
[10] A. Crewe, D. N. Eggenberger, J. Wall, L. M. Welter, *Rev. Sci. Instrum.* **1968,** *39*, 576.
[11] J. Munch, *Optik* **1975,** *43*, 79.
[12] J. A. Lin, J. M. Cowley, *Ultramicroscopy* **1986,** *19*, 31.
[13] J. Rodenberg, R. Bates, *Phil. Trans. R. Soc., Ser. A* **1992,** *339*, 521.
[14] H. W. Fink, H. Schmid, H. J. Kreuzer, A. Wierzbicki, *Phys. Rev. Lett.* **1991,** *67*, 1543.
[15] H. Fink, W. Stocker, H. Schmid, *Phys. Rev. Lett.* **1990,** *65*, 1204.
[16] H. Fink, W. Stocker, H. Schmid, *J. Vac. Sci. Technol. B* **1990,** *8*, 1323.
[17] R. Morin, A. Gargani, *Phys. Rev. B* **1993,** *48*, 6643.
[18] J. C. H. Spence, W. Qian, A. J. Melmed, *Ultramicroscopy* **1993,** *52*, 473.
[19] V. T. Binh, V. Semet, N. Garcia, *Ultramicroscopy* **1995,** *58*, 307.
[20] M. Scheinfein, W. Qian, J. Spence, *J. Appl. Phys.* **1993,** *73*, 2057.
[21] D. K. Saldin, J. C. H. Spence, *Ultramicroscopy* **1994,** *55*, 397.
[22] D. R. Penn, *Phys. Rev. B* **1987,** *35*, 482.
[23] M. Seah, W. Dench, *Surf. Interface Anal.* **1979,** *1*, 2.
[24] S. Tanuma, C. Powell, D. Penn, *Surf. Interface Anal.* **1991,** *17*, 911.
[25] C. Martin, E. Arakawa, T. Callcott, R. Warmack, *J. Electron Spectrosc. Rel. Phenom.* **1987,** *42*, 171.
[26] W. Qian, J. Spence, J. Zuo, *Acta Crystallogr., Part A* **1993,** *49*, 436.
[27] J. Kreuzer, K. Nakamura, A. Wierzbicki, H. Fink, H. Schmid, *Ultramicroscopy* **1992,** *45*, 381.
[28] J. C. H. Spence, W. Qian, *Phys. Rev. B* **1992,** *45*, 10271.
[29] G. M. Shedd, *J. Vac. Sci. Technol. A* **1994,** *12*, 2595.
[30] J. C. H. Spence, *Optik* **1992,** *92*, 57.
[31] J. DeVelis, G. Parrent, B. Thompson, *J. Opt. Soc. Am.* **1966,** *56*, 423, and references therein.
[32] J. Spence, *Experimental High Resolution Electron Microscopy,* Oxford University Press, Oxford, 1988.
[33] J. C. H. Spence, X. Zhang, W. Qian, *Proc. Int. Workshop on Electron Holography* **1995** (Eds.: A. Tonomura, L. Allard, D. Joy, Y. Ono), Elsevier, Amsterdam, 267.
[34] J. C. H. Spence, W. Qian, M. P. Silverman, *J. Vac. Sci. Technol. A* **1994,** *12*, 542.
[35] A. P. Janssen, J. P. Jones, *J. Phys. D: Appl. Phys.* **1971,** *4*, 118. See also: D. Bettler, S. Charbonnier, *Phys. Rev.* **1960,** *119*, 85.
[36] Vu Thien Binh, S. T. Purcell, N. Garcia, J. P. Doglioni, *Phys. Rev. Lett.* **1992,** *69*, 2527.
[37] Vu Thien Binh, N. Garcia, K. Dransfeld (Eds.), *Manipulation of Atoms under High Fields and Temperatures: Applications,* NATO-ASI-E Series, Vol. E235, Plenum Press, New York **1993**.
[38] A. Degiovanni, R. Morin, *Proc. 13th Int. Congress on Electron Microscopy*, Les Editions de Physique (Les Ulis), **1994,** Vol. I, p. 331.
[39] W. Qian, M. Scheinfein, J. Spence, *J. Appl. Phys.* **1993,** *73*, 7041.
[40] P. Bovet, Personal communication.
[41] J. M. Cowley, *Ultramicroscopy* **1979,** *4*, 435.
[42] P. Chen, *Introduction to Scanning Tunnelling Microscopy*, Oxford University Press, Oxford **1993**.
[43] J. C. H. Spence, *Proc. Int. Congress on Electron Microscopy* Les Editions de Physique (Les Ulis) **1994,** 329.
[44] L. Howald, H. Rudin, H. Guntherodt, *Rev. Sci. Instrum.* **1992,** *63*, 3909.
[45] A. J. Melmed, *J. Vac. Sci. Technol. B* **1991,** *9*, 601.
[46] J. C. H. Spence, J. M. Zuo, *Electron Microdiffraction*, Plenum Press, New York **1992,** Fig. 8.11.

[47] J. C. H. Spence, J. M. Cowley, *Optik* **1978,** *50*, 129.
[48] J. Rodenberg, B. McCallum, P. Nellist, *Ultramicroscopy* **1993,** *48*, 304.
[49] J. M. Cowley, *J. Electron. Microsp. Technol.* **1986,** *3*, 25.
[50] H. Schmid, H. W. Fink, *Appl. Surf. Sci.* **1993,** *67*, 436 and references therein.
[51] V. E. Cosslett, W. C. Nixon, *J. Appl. Phys.* **1953,** *24*, 616.
[52] J. Clauser, M. Reinsch, *Appl. Phys. B* **1992,** *54*, 380.
[53] A. Howie, M. Muhid, F. Rocca, U. Valdre, *Inst. Phys. Conf. Ser. 90 (EMAG87)* **1987,** 155.
[54] J. Spence, W. Qian, X. Zhang, *Ultramicroscopy* **1994,** *55*, 19.
[55] X. Zhang, J. Spence, *Proc. 53rd Annual Meeting of the Electron Microscopy Society of America* (Ed.: G. Bailey), San Francisco Press, San Francisco **1995,** p. 846.
[56] M. Isaacson, D. Johnson, A. V. Crewe, *Rad. Res.* **1973,** *55*, 205.
[57] A. J. Swallow, *Radiation Chemistry*, Longmans–Halsted–Wiley, New York **1973.**
[58] R. Henderson, Quarterly Reviews of Biophysics **1995,** *28*, 171.
[59] J. Spence, W. Qian, *Proc. 50th Annual Meeting of the Electron Microscopy Society of America* (Ed.: G. Bailey), San Francisco Press, San Francisco **1992,** p. 938.
[60] J. Spence, W. Qian, W. Lo, S. Mo, U. Knipping, X. Zhang, in *Determining Nanoscale Physical Properties of Materials by Microscopy and Spectroscopy* (Ed.: M. Sarikaya) MRS, Pittsburgh **1994,** Vol. *332*, 405.
[61] J. C. H. Spence, X. Zhang, J. M. Zuo, U. Weierstall, E. Munro, J. Rouse, in *Proc. MSA* (Ed.: G. Bailey), San Francisco Press, San Francisco **1995,** 610.
[62] J. C. H. Spence, X. Zhang, J. Zuo, U. Weierstall. *Surf. Rev. Lett.* **1996,** in press.
[63] H. Rose, *Optik* **1987,** *77*, 26.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

General Reading

Chapter IV.2.1

J. N. Chapman and A. J. Craven (Ed.), *Quantitative Electron Microscopy*, Scottish University Summer Schools in Physics, **1984**.

J. I. Goldstein, D. E. Newbury, P. E. Echlin, D. C. Joy, A. D. Romig, C. E. Lyman, C. E. Fiori, E. Lifshin, *Scanning Electron Microscopy and X-Ray Microanalysis*, Plenum Press, New York, **1992**.

D. B. Holt and D. C. Joy, *SEM Microcharacterization of Semiconductors*, Academic Press, London, **1989**.

C. E. Lyman, D. E. Newbury, J. I. Goldstein, D.B. Williams, A. D. Romig, J. T. Armstrong, P. Echlin, D. C. Joy, E. Lifshin and K. R. Peters, *SEM, X-Ray Microanalysis and AEM – A Laboratory Notebook*, Plenum Press, NY, **1990**.

D. E. Newbury, D. C. Joy, P. Echlin, C. E. Fiori and J. I. Goldstein, *Advanced Scanning Electron Microscopy and X-Ray Microanalysis*, Plenum Press, NY, **1986**.

C. W. Oatley, *The early history of the SEM*, J. Appl. Phys., 53, R1, **1982**.

L. Reimer, *Image Formation in Low Voltage SEM*, SPIE Tutorial Texts in Optical Engineering, (SPIE: Bellingham, WA), TT12, **1993**.

L. Reimer, *Scanning Electron Microscopy Physics of Image Formation and Microanalysis*, Springer Series in Opt. 5a. 45, Springer, Berlin, **1985**.

O. C. Wells, *Scanning Electron Microscopy*, McGraw-Hill, NY, **1974**.

B. G. Yacobi and D. B. Holt, *Cathodoluminescence of Inorganic Solids*, Plenum Press, NY, **1990**.

Chapter IV.2.2

J. M. Cowley, *Special Electron Microscopy Techniques*, Chapt. 7, in *Physical Methods of Chemistry*, Vol. IV, *Microscopy*, (Eds. B. W. Rossiter, J. F. Hamilton), John Wiley & Sons, New York, **1991**, p. 239.

A. V. Crewe, *The physics of the high-resolution scanning microscope, Rep. Prog. Phys.*, **1980**, *43*, 621.

H. Rose, *Non-standard imaging methods in electron microscopy, Ultramicroscopy,* **1977**, *2*, 251.

M. Isaacson, D. Kopf, M. Utlaut, N. W. Parker, A. V. Crewe, *Direct observations of atomic diffusion by scanning transmission electron microscopy,* Proc. Natl. Acad. Sci. USA, **1977**, *74*, 1802.

J. C. H. Spence and J. M. Zuo, *Electron Microdiffraction*, Plenum Press, New York, **1992**.

J. M. Cowley, *Electron nanodiffraction; progress and prospects, J. Electron Microscopy,* **1996**, (in press).

J. M. Cowley, J. C. H. Spence, *Principles and theory of electron holography*, Chapter 1 of *Introduction to Electron Holography*, L. F. Allard et al. (Eds.), Plenum Press, New York, **1996** (in press).

S. J. Pennycook, D. E. Jesson, *High-resolution Z-contrast imaging of crystals. Ultramicroscopy,* **1991**, *37*, 14.

J. M. Cowley, *Scanning microscopy at Arizona State University.* In: EMSA Bulletin, **1991**, *21*, 57.

J. Liu, J. M. Cowley, *Scanning reflection electron microscopy and associated techniques for surface studies. Ultramicroscopy,* **1993**, *48*, 381.

Chapter IV.2.3

D. M. Bird, *Theory of Zone Axis Electron Diffraction*, J. Elect. Micr. Tech. *13*, **1989**, 77.

P. R. Buseck, J. M. Cowley and L. Eyring (Eds.), *High Resolution Transmission Electron Microscopy*, Oxford University Press, New York, **1988**.

J. N. Chapman and A. J. Craven (Eds.), *Quantitative Electron Microscopy*, Scottish University Summer School in Physics, Edinburgh, **1984**.

M. M. Disko, C. C. Ahn and B. Fultz (Eds.), *Transmission Electron Energy Loss Spectrometry in Materials Science*, The Minerals, Metals & Materials Society, Warrendale PA, **1992**.

R. F. Egerton, *Electron Energy-Loss Spectroscopy in the Electron Microscope*, Plenum Press New York, **1986**.

C. J. Humphreys, *The Scattering of Fast Electrons by Crystals*, Rep. Prog. Phys., *42*, **1979**, 1825.
P. G. Merli and M. Vittori (Eds.), *Electron Microscopy in Materials Science*, Antisari, World Scientific, Singapore, **1992**.
S. J. Pennycook, *High Resolution Electron Microscopy and Microanalysis*, Contemporary Physics *23*, **1982**, 371.
S. J. Pennycook, *Z-Contrast Imaging in the Scanning Transmission Electron Microscope*, J. Microsc. Soc. Am. *1*, **1995**, 231.
L. Reimer, *Transmission Electron Microscopy*, Springer-Verlag, Berlin, **1993**.
J. Wall, J. Langmore, M. Isaacson and A. V. Crewe, *Scanning Transmission Electron Microscopy at High Resolution*, Proc. Nat. Acad. Sci. USA, *71*, **1974**, 1.

Chapter IV.2.4

G. C. Smith, *Surface Analysis by Electron Spectroscopy*, Plenum Press, New York and London, **1994**.
D. Briggs and M. P. Seah, *Practical Surface Analysis*, Second Edition, Vol. 1, John Wiley and Sons, **1990**.
L. C. Feldman and J. W. Mayer, *Fundamentals of Surface and thin film Analysis*, North Holland Elsevier Science Publishing Co. Inc. **1986**.
W. F. Egelhoff, Jr., *Surface Science Reports 6*, **1987**, 253-415.
A. W. Czanderna, *Methods of Surface Analysis*, Elsevier Scientific Publishing Company, Amsterdam, **1975**.
W. Dekeyser, L. Fiermans, G. Vanderkelen and J. Vennik, *Electron Emission Spectroscopy*, D. Reidel Publ. Co., Dordrecht, **1973**.
L. Fiermans, J. Vennik and W. Dekeyser, *Electron and Ion Spectroscopy of Solids*, Plenum Press, New York, **1978**.
T. A. Carlson, *Photoelectron and Auger spectroscopy*, Plenum Press, New York, **1975**.

Chapter IV.2.5

M. A. Baker and J. E. Castle, in: *Material Science and Technology: a Comprehensive Treatment*, (Ed. R. W. Cahn, P. Haasen and E. I. Kramer), VCH, Weinheim, **1994**, Vol. 2B, Chapter 13.
N. D. Browning and S. J. Pennycook, *Microbeam Analysis*, **1993**, *1*, 19.
J. I. Goldstein, D. E. Newbury, P. Echlin, D. C. Joy, C. E. Fiori and E. Lifshin, *Scanning Electron Microscopy and X-Ray Microanalysis*, Plenum Press, New York, **1981**.
J. I. Goldstein and D. B. Williams, *Microbeam Analysis*, **1992**, *1*, 29.
J. J. Hren, J. I. Goldstein and D. C. Joy (Ed.), *Introduction to Analytical Electron Microscopy*, Plenum Press, New York, **1979**.
D. C. Joy, *Microbeam Analysis*, **1992**, *1*, 19.
D. C. Joy, A. D. Romig and J. I. Goldstein (Ed.), *Principle of Analytical Electron Microscopy*, Plenum Press, New York, **1986**.
L. Reimer, *Transmission Electron Microscopy. Physics of Image Formation and Microanalysis*, 2nd ed., Springer- Verlag, Berlin, **1989**.
K. F. J. Heinrich and D. E. Newbury (Ed.), *Electron Probe Quantification*. Plenum Press, New York, **1991**.
V. D. Scott, G. Love and S. J. B. Reed, *Quantitative Electron-Probe Microanalysis*, 2nd ed., Ellis Horwood, New York, **1995**.

Chapter IV.2.6.

A. Benninghoven, F. G. Rüdenauer, H. W. Werner (Eds.), *Secondary Ion Mass Spectrometry: Basis Concepts, Instrumental Aspects, Applications and Trends*, Wiley, New York, **1987**.
A. Benninghoven, *Angew. Chem.* Int. Ed. Engl., **1994**, *33*, 1023.
K. T. F. Janssen, P. R. Boudewijn, in: *Analysis of Microelectronic Materials and Devices* (Eds.: M. Grasserbauer, H. W. Werner), Wiley, Chichester, **1991**, p. 407.
A. Lodding, in: *Inorganic Mass Spectrometry* (Eds.: F. Adams, R. Gijbels, R. Van Grieken) Wiley, New York, **1988**, p. 125.
R. W. Odom, in: *Microscopic and Spectroscopic Imaging of the Chemical State* (Ed.: M. D. Morris), Marcel Dekker, New York, **1993**, p. 345.
J. C. Vickerman, A. Brown, N. M. Reed (Eds.), *Secondary Ion Mass Spectrometry: Principles and Applications*, Charendon Press, Oxford, **1989**.
R. G. Wilson, F. A. Stevie, C. W. Magee, *Secondary Ion Mass Spectrometry*, Wiley, New York, **1989**.

Chapter V.1

B. Blümich, W. Kuhn, *Magnetic Resonance Microscopy*, VCH, Weinheim, **1992**.

P. Blümler, B. Blümlich, *NMR-Basic Principles and Progress*, Springer-Verlag, Berlin, **1993**.

P. T. Callaghan, *Principles of Nuclear Magnetic Resonance Microscopy*, Oxford University Press, Oxford, **1991**.

D. G. Cory, *Annu. Rep. NMR Spectrosc.*, **1992**, *24*, 88.

P. Jezzard, J. J. Attard, T. A. Carpenter, L. D. Hall, *Prog. NMR Spectrosc.*, **1992**, 23.

P. Mansfield, E. L. Hahn, *NMR Imaging*, The Royal Society, London, **1990**.

Chapter V.2

R. J. Celotta, D. T. Pierce and J. Unguris, *MRS Bulletin*, **1995**, 30.

E. D. Dahlberg and J. G. Zhu, *Physics Today*, **1995**, 34.

J. P. Jacubovics, *Electron Microscopy in Materials Science Part IV*, Commission of European Communities, Brussels, **1973**, p. 1305.

J. Kessler, *Polarized Electrons*, 2nd ed., Springer-Verlag, Berlin, **1985**.

J. Kirschner, *Polarized Electrons at Surfaces*, Springer-Verlag, Berlin, **1985**.

D. E. Newbury, D. E. Joy, P. Echlin, C. E. Fiori, J. I. Goldstein, *Advanced Scanning Electron Microscopy and X-Ray Microanalysis*, Plenum Press, New York, **1986**.

D. T. Pierce, J. Unguris, R. J. Celotta, in *Ultrathin Magnetic Structure II*, Springer-Verlag, Berlin, **1994**, p. 117.

J. Unguris, M. R. Scheinfein, R. J. Celotta, D.T. Pierce, in *Chemistry and Physics of Solid Surfaces VII*, Springer-Verlag, Germany, **1990** p. 239.

Chapter V.3

E. Bauer, *Rep. Prog. Phys.*, **1994**, *57*, 895.

Chapter VI.1

E. Bauer, W. Telieps, in: *Surface and Interface Characterization by Optical Methods* (Eds.: A. Howie and U. Valdré), Plenum, New York, **1988**, p. 195.

E. F. Burton, W. H. Kohl, *The Electron Microscope*, 2nd ed., Reinhold, New York, **1948**.

H. Bethge, J. Heydenreich, *Electron Microscopy in Solid State Physics*, Elsevier, Amsterdam, **1987**.

V. E. Cosslett, *Introduction to Electron Optics*, 2nd ed., Oxford University Press, Oxford, **1950**.

O. H. Griffith, W. Engel, *Ultramicroscopy*, **1991**, *36*, 1.

O. H. Griffith, G. F. Rempfer, *Advances in Optical and Electron Microscopy*, Vol. 10, **1987**, p. 269.

C. E. Hall, *Introduction to Electron Microscopy*, McGraw-Hill, New York, **1953**.

O. Klemperer, *Electron Optics*, 2nd ed., Cambridge University Press, Cambridge, **1953**.

C. Ramsauer, *Electronenmikroskopie (Report of the Work of the AEG Research Institute 1930-1942)*, Springer, Berlin, **1943**.

R. A. Schwarzer, *Microsc. Acta*, **1981**, *84*, 51.

Chapter VI.2

A. Cerezo, G. D. W. Smith, *Materials Science and Technology*, Vol. 2B: Characterisation of Materials Part II, (Ed. E. Lifshin), VCH, Weinheim, pp. 513-559.

Chapter VII.1

R. Wiesendanger, *Scanning Probe Microscopy and Spectroscopy: Methods and Applications*, Cambridge University Press, Cambridge, **1994**.

C. J. Chen, *Introduction to Scanning Tunneling Microscopy*, Oxford University Press, New York and Oxford, **1993**.

H.-J. Güntherodt and R. Wiesendanger (Eds.), *Scanning Tunneling Microscopy I*, Springer, Berlin, Heidelberg, New York, 2nd edition, **1994**.

R. Wiesendanger and H.-J.Güntherodt (Eds.), *Scanning Tunneling Microscopy II*, Springer, Berlin, Heidelberg, New York, 2nd edition, **1995**.

R. Wiesendanger and H.-J. Güntherodt (Eds.), *Scanning Tunneling Microscopy III*, Springer, Berlin, Heidelberg, New York, 2nd edition, **1996**.

J. Stroscio and W. Kaiser (Eds.), *Scanning Tunneling Microscopy*, Academic Press, **1993**.

D. A. Bonnel (Ed.), *Scanning Tunneling Microscopy and Spectroscopy: Theory, Techniques and Applications*, VCH-Verlag, **1993**.
S. Gauthier and C. Joachim (Eds.), *Scanning Probe Microscopy: Beyond the Images*, Les éditions de physique, **1992**.
P. Avouris (Ed.), *Atomic and Nanometer-Scale Modification of Materials: Fundamentals and Applications*, Kluwer, **1993**.
C. R. K. Marrian (Ed.), *Technology of Proximal Probe Lithography*, SPIE Press, **1993**.

Chapter VII.2

H.-J. Güntherodt, D. Anselmetti and E. Meyer (Eds.): *Forces in Scanning Probe Methods*, NATO-ASI Series, Series E: Applied Sciences, Vol. 286, Kluwer Academic Publishers, Dordrecht, **1995**.

Chapter VII.3

Dror Sarid, *Scanning Force Microscopy*, Oxford University Press, Oxford, New York, **1991**.
H.-J. Güntherodt, D. Anselmetti and E. Meyer (Eds.), *Forces in Scanning Probe Methods*, Kluwer Academic Publishers, Dordrecht, Boston, London, **1995**.
Sushin Chikazumi, *Physics of Magnetism*, John Wiley & Sons, Inc., **1964**.
Richard M. Bozorth, *Ferromagnetism*, D. van Nostrand Comp., Inc., New York, **1951**.
W. F. Brown, Jr. Robert, E. Krieger, *Micromagnetics*, Publish. Comp., Huntington, New York, **1978**.
D. J. Craik and R. S. Tebble, *Ferromagnetism and Ferromagnetive Domains*, North Holland Publishing Comp., Amsterdam, **1965**.
A. P. Malozemott and J. C. Slonczewski, *Magnetic Domain Walls in Bubble Materials*, Academic Press, INC., New York, **1979**.

Chapter VIII.1

J. C. Dainty and R. Show, *Image Science*, Academic Press, London, **1974**.
P. N. J. Dennis, *Photodetectors: An introduction to current technology*, Plenum Press, New York, **1985**.
L. M. Biberman, S. Nudelman (Eds.), *Photoelectronic Imaging Devices, Vol. I and II*, Plenum Press, New York, **1971**.
M. M. Blouke (Ed.), *Charge-coupled Devices and Solid State Optical Sensors IV*, SPIE Proc., **1994**, 2172.
Photometrics Ltd., *Charge-coupled Devices for Quantitative Electronic Imaging*, Photometrics Ltd., Tucson, **1989**.
D. Shotton (Ed.), *Electronic light microscopy, Techniques in Modern Biomedical Microscopy*, Wiley-Liss, New York, **1993**.
K.-H. Herrmann and D. Krahl, *Electronic image recording in conventional electron microscopy*, Adv. Opt. El. Micr. (Ed. R. Barer, V. E. Cosslett), **1984**, 9, 1.
W. J. de Ruijter, *Imaging properties and applications of slow-scan charge-coupled device cameras suitable for electron microscopy*, Micron, *26*, **1995**, 247.
G. Schmahl and D. Rudolph (Eds.), *X-Ray Microscopy*, Springer Series in Optical Sciences, **1984**, 43.
A. G. Michette, G. R. Morrison and C. J. Buckley (Eds.), *X-Ray Microscopy III*, Springer Series in Optical Sciences, **1992**, 67.

Chapter IX.2

J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, Current Edition, New York.
J. C. H. Spence and J. M. Zuo, *Electron Microdiffraction*, Plenum Press, New York, **1992**, (ISBN 0-306-44262-0).
J. C. H. Spenc, *Experimental High Resolution electron Microscopy*, Oxford University Press, New York, **1988**, 2nd. Edn. (ISBN 0-19-505405-9).
R. Gomer, *Field Emission and field Ionization*, American Institute of Physics, New York, **1993**.
R. Collier, C. Burkhadt, L. Lin, *Optical Holography*, Academic Press, New York, **1971**.

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

List of Symbols and Abbreviations

Symbol	Designation	Chapter
A	amplitude	VII.4
A	area	IV.2.6, VIII.2, VIII.1
A	atomic weight of element a; constant	IV.2.5
A	fitting parameter	IX.2
A	scattering asymmetry	V.2
a	constant	V.1
a	periodicity	IV.2.2
$A\,(u)$	aperture function	IV.2.2
A, B	linear dimensions	IV.2.1
a, b	constant	VI.2
A_C	coherence area	IX.2
A_{ex}	exchange asymmetry	V.3
a_{ext}	amplitude of vibration of support	IX.2
a_{tip}	amplitude of vibration of tip	IX.2
B	fitting parameter; brightness	IX.2
$\boldsymbol{B}$	magnetic field	V.1, IX.2
$\boldsymbol{B}$	magnetic flux density	V.3
b	Kramers' constant (2×10^{-9} photons s^{-1} eV^{-1} $electron^{-1}$)	IV.2.5
B_0	upper quantum brightness	IX.2
$b^j(\boldsymbol{K}, \boldsymbol{R})$	Bloch state amplitude	IV.2.3
b^{RMS}	Root-mean-square impact parameter for inelastic scattering	IV.2.3
C	complex constant	IX.2
C	constant	IV.2.5
C	image contrast	VIII.1
c	spring constant of the cantilever	VII.3
c	velocity of light	VIII.1
C(x, y)	contrast	VIII.2
C, C'	local contrast	VIII.2
c_0	estimated concentration	VI.2
C_1, C_2, C_3	constants	VI.2
C_c	Chromatic aberration coefficient	IX.2
C_i, C_j	concentration of element i, j	IV.2.5
C_a^j	mass concentration of element a in all elements j	IV.2.5
C_M	concentration of species $\boldsymbol{M}$	IV.2.6
C_s	Spherical aberration coefficient	IX.2, IV.2.3, IV.2.1, IV.2.2
c_s	speed constant of a photoemulsion	VIII.1
C_X	atomic concentration of element X	IV.2.4
D	fractal dimension	VIII.2

Symbol	Designation	Chapter
D	optical density; charge-coupled device dark current ($e^- pixel^{-1}s^{-1}$)	VIII.1
D	pinhole spacing	IX.2
D	specimen–screen distance	VI.2
d	crystal lattice spacing	IV.2.4
d	diameter of electron beam probe	IV.2.1
d	distance between image planes	VII.2
d	distance between pixels	VIII.2
d	flight length	VI.2
d	fringe spacing; tip size; virtual source size; width of object	IX.2
d	lattice spacing	IV.2.5
d	object size to be resolved	VIII.1
$D\,(E, T)$	total transmission probability for electron transport to the back of the semiconductor	VII.4
$D\,(\boldsymbol{K})$	Reciprocal space annular detector function	IV.2.3
$d\,(\boldsymbol{R})$	Real space annular detector function	IV.2.3
$D\,(u)$	transmission function of detector	IV.2.2
$d_{(hkl)}$	interplanar spacing for (hkl) crystallographic pole	VI.2
$d(\hat{o}^{(k)})$	difference wave	VIII.2
D_0	fog density of photoplates	VIII.1
d_0	thickness of an absorbing layer	VIII.1
D_A	detector efficiency	IV.2.4
d_{CCD}	pixel size of a charge-coupled device	VIII.1
d_d	thickness of a dead layer	VIII.1
D_s	saturation density of photoplates	VIII.1
D_X	relative scale factor	IV.2.4
E	electron energy	VII.4, IX.2
$\boldsymbol{E}$	energy	IV.2.5, V.1, VII.2, VIII.1
$\mathscr{E}$	energy with respect to the Fermi level	VII.2
E	incident electron energy	IV.2.3
E	mean grey level of edges; expectation value	VIII.2
E	tip-sample interaction energy	VII.4
e	electron charge	IV.2.2, IV.2.3, VI.2, VII.2, VII.4, VIII.1, IX.1, IX.2
$E(z)$	on-axis electric field	IX.2
E_0	energy of a primary electron	VIII.1, IV.2.5
E_0	incident electron beam energy	IV.2.1
E_A	Auger kinetic energy	IV.2.4
E_{abs}	volume density of absorbed energy	VIII.1
E_b	ground state binding energies	IV.2.4
$\boldsymbol{E}_c$	energy of critical excitation band	IV.2.5
E_d	material „dead voltage“	IV.2.1
E_{eh}	energy required to form a one-hole electron pair	IV.2.1
E_F	Fermi energy	IV.2.5, VII.2, VII.4
E_f	mean formation energy of an electron-hole pair	VIII.1
E_{gap}	material band-gap energy electron	IV.2.1

Symbol	Designation	Chapter
E_{in}	spatial density of energy	VIII.1
E_K, E_L, ...	energy of the Kth and Lth shells	IV.2.4
E_{kin}	kinetic energy	IV.2.4
E_μ	Energy of the state Ψ_m	VII.2
E_v	Energy of the state Ψ_n	VII.2
E_{ph}	photon energy	VIII.1
$E_{ph,\,x}$	energy of an X-ray photon	VIII.1
eu	incident electron energy (eV)	IV.2.5
F	electric field strength	VI.2
F	F-number of a lense	VIII.1
F	figure of merit	V.2
F	force	VII.3
F	fraction of total scattering passing through the hole in the annular detector	IV.2.3
F	transverse electric field	IX.2
F'	overall force gradient	VII.4
f	focal length of a lense	VIII.1
f	function	IV.2.5
$f(E, T)$	energy distribution of electrons reaching the interface	VII.4
$f(\boldsymbol{K}), f(s)$	atomic form factor for elastic scattering	IV.2.3
$f(u)$	atomic scattering factor	IV.2.2
$f(X)$	fraction of carrier which diffuse to junction	IV.2.1
$f(\chi)$	absorption factor	IV.2.5
$f(E)$	Fermi function	VII.2
F_a	Fano factor	VIII.1
F_e	electrical field strength	VIII.1
F_i	fluorescence correction for element i	IV.2.5
F_z	force acting along the z-axis	VII.4
f_{res}	resonance frequency of the cantilever	VII.3
f_{sp}	resonance frequency of spring	IX.2
f_{tip}	resonance frequency of tip	IX.2
G	Fourier component	VII.2
G	spectral matching factor for a phosphor sensor	VIII.1
g	normalization factor	IV.2.5
$G(\boldsymbol{u}, \rho)$	a four-dimensional function	IV.2.2
G(x, y)	gray level gradient magnitude	VIII.2
g_m	quantum yield of a converting stage	VIII.1
H	transfer function; parameter	VIII.2
h	constant	VII.4
h	impulse response	VIII.2
h	Planck's constant	IV.2.1, IV.2.4, VII.2, VIII.1, IV.2.2
$\hbar$	Dirac's constant	V.1, VII.2
$\hbar$	Planck's constant	IV.2.3
H(r)	magnetic field above a sample	VII.4
$H(x^{-1}, y^{-1})$	transfer function of an imaging instrument	VIII.2
H_x	stray magnetic force along the x-axis	VII.4
H_z	stray magnetic force along the z-axis	VII.4
h(x, y)	impulse response of an imaging instrument	VIII.2

Symbol	Designation	Chapter
I	current	VII.4, VI.2
I	current of primary electron beam	IX.1
I	Fourier transform of i	VIII.2
I	nuclear spin	V.1
I	reflected intensity	V.3
$\boldsymbol{I}$	spin angular momentum	V.1
I	tunneling current	VII.2
i	image function	VIII.2
$I(\Theta)$	angular distribution of backscattered current in Eq. 3	V.2
$I(\boldsymbol{R}_0)$	Image intensity for probe in position R_0	IV.2.3
$I(\boldsymbol{R}_0, t)$	Thickness dependent image intensity incorporating dynamical effects	IV.2.3
$I(x)$	intrensity of hologram	IX.2
i(x, y)	image function	VIII.2
$I(z)$	Intensity of electron beam as a function of depth within sample	IX.2
I, I'	pixel grey level	VIII.2
$I_M^{\pm}$	secondary in current of species $\boldsymbol{M}$	IV.2.6
I_0	current of primary electron beam	IV.2.4
I_0	Intensity of electron beam incident on sample	IX.2
I_0	polarization-independent intensity	V.3
I_{00}	incident intensity	V.3
I_A	Auger yield	IV.2.4
I_A/I_B	Auger current ratio for elements A and B	IV.2.4
I_{Ag}	peak-to-peak amplitude of silver standard	IV.2.4
I_B	current required in the beam	IV.2.1
I_b	ballistic electron emission microscopy current	VII.4
I_b	current contained in electron-beam probe	IV.2.1
I_c	collector current	VII.4
I_{cc}	charge collected signal	IV.2.1
I_{CL}	brightness dependence of cathodoluminescence signal	IV.2.1
I_{ex}	exchange intensity	V.3
$I^H(\boldsymbol{R}_0)$	Image intensity using a high angle annular detector	IV.2.3
I_i	current induced in segment i of a photodiode	VII.3
I_{sp}^i	intensity of line i in specimen	IV.2.5
I_{st}^i	intensity of line i in standard	IV.2.5
I_i, I_j	X-ray intensity of line i, j	IV.2.5
I_k	kth ionization energy	VI.2
I_P	primary in current	IV.2.6
I_s	signal current of a camera tube	VIII.1
I_t	injection current	VII.4
I_X	peak-to-peak amplitude of element X	IV.2.4
$I_X(u)$	intensity distribution	IV.2.2
j	current density	IX.2
j	electron current density	VIII.1
j	inner quantum number	IV.2.5
$J(\boldsymbol{r}\text{-}\boldsymbol{r}_i)$	exchange coupling strength	V.3
j, k, l	energy levels relating to multiplet splitting and electronic relaxation	IV.2.4
$j_{\mu,\nu}$	current density between the states Ψ_m and Ψ_n	VII.2
K	attenuation of vibration	IX.2
K	instrumental constant	IV.2.4

Symbol	Designation	Chapter
$\boldsymbol{K}$	transverse component of the incident wavevector χ	IV.2.3
k	Boltzmann's constant	V.1, VIII.1, VII.4
k	constant relating electric field at specimed apex to the applied voltage	VI.2
$\boldsymbol{k}$	electron wave vector	V.3
$\boldsymbol{k}$	scattering vector	IV.2.3
k	spring constant	VII.4
K, L, M, ...	atomic energy shells	IV.2.4
$k_{\|\|}$	electron momentum parallel to interface	VII.4
k_B	Boltzmann's constant	VI.2
k_β	Boltzmann's constant	VII.2
$\boldsymbol{K}_c$	cut-off wavevector	IV.2.3
$\boldsymbol{K}_f$	transverse component of the final state wavevector	IV.2.3
k_{eff}	effective spring constant	VII.4
k_a^i	I_{sp}^i / I_{st}^i ratio for element a	IV.2.5
$\boldsymbol{k}_i$, $\boldsymbol{k}_f$	incident and scattered wave vectors, respectively	V.2
k_{ij}	Cliff-Lorimer sensitivity factor	IV.2.5
k_z, k_A, k_F	atomic number, absorption, fluorescence correction factor	IV.2.5
L	attenuation length; supercell period	IX.2
L	minority carrier diffusion length	IV.2.1
$\boldsymbol{L}$	orbital angular momentum	V.2
L	total length of boundaries	VIII.2
l	azimuthal quantum number	IV.2.5
L_c	longitudinal coherence length	IX.2
l_c	coherence length	IV.2.3
M	atomic mass	IV.2.6
M	Debye-Waller factor	IV.2.3
M	instrumental magnification; linear magnification	IV.2.1
$\boldsymbol{M}$	magnetic moment	V.1
$\boldsymbol{M}$	magnetization	V.3
M	magnification	IX.2
M	scaling factor, molecular mass	VIII.1
M	tip magnetization	VII.4
m	cantilever mass	VII.4
m	free electron mass	VII.4
m	mass	IV.2.3, VII.2, IV.2.2, VI.2
$\overline{m}_s$	electron effective mass in a semiconductor	VII.4
$M_{\mu, n}$	tunneling matrix element between the states Ψ_m and Ψ_n	VII.2
M_x, M_y, M_z	components of magnetization M	V.2
N	Avogadro's number	IV.2.5
N	density of atoms	IX.2
N	Fourier transform of n	VIII.2
N	number of primary quanta	VIII.1
N	total number of atoms; number of image rings	VI.2
n	additive noise function	VIII.2
n	atomic density	IX.1
n	constant	IV.2.1
n	diffraction order	IV.2.4

Symbol	Designation	Chapter
n	number of atoms in a column	IV.2.3
n	number of electrons removed during ionisation of an atom	VI.2
n	number of occupied d orbitals	V.3
n	principal quantum number; Bragg constant	IV.2.5
n	spatial density of primary quanta	VIII.1
$\boldsymbol{n}$	unit vector normal to the scattering plane	V.2
$\bar{n}$	average number of scattering events	VII.4
N	number of electrons detected with spins parallel and antiparallel to x direction	V.2
n	number of spins per unit volume parallel and antiparallel to a particular orientation	V.2
$N(d)$	Fresnel number	IX.2
$N(E)$	distribution of incident particles	IV.2.4
$N(E)\ \mathrm{d}(E)$	number of photons with energy in the range E to $E + \mathrm{d}E$	IV.2.5
n_0	refractive index of a scintillator	VIII.1
n_1	refractive index of fiber-optic core	VIII.1
n_2	refractive index of a fiber-optic coat	VIII.1
N_A	analyte concentration	IV.2.4
NA	numerical aperture	VIII.1
N_A, N_B	number of points inside structures A, B	VIII.2
n_B	number of Bohr magnetrons per atom	V.2
n_e	electron response of a single primary quantum	VIII.1
N_L, N_R	number of electrons scattered to left and right, respectively	V.2
n_{max}	full-well capacity of a charge-coupled device (e^- pixel^{-1})	VIII.1
N_p	number of primary quanta on a pixel	VIII.1
n_{ph}	photon response of a single primary quantum	VIII.1
n_r	read-out noise (e^- pixel^{-1})	VIII.1
n_s	number of atoms of species s	VI.2
n_s	number of secondaries per primary quantum	VIII.1
$n_s(E)$	density of states for the sample	VII.2
$n_t(E)$	density of states for the tip	VII.2
n_v	number of valence electrons per atom	V.2
O	Fourier transform of o	VIII.2
o	object function	VIII.2
$O'(\boldsymbol{K})$	Reciprocal space object function for inelastically scattered electrons	IV.2.3
$O'(\boldsymbol{R})$	Real space object function for inelastically scattered electrons	IV.2.3
$O(\boldsymbol{K})$	Reciprocal space object function for elastically scattered electrons	IV.2.3
$O(\boldsymbol{R})$	real space object function for elastically scattered electrons	IV.2.3
$O(\boldsymbol{R}, t)$	thickness dependent object function incorporating dynamical effects	IV.2.3
$O(x^{-1}, y^{-1})$	object spectrum = Fourier transform of the object function	VIII.2
$o(x, y)$	object function	VIII.2
$O^B(\boldsymbol{R})$	Born approximation object function	IV.2.3
$O^{TDS}(\boldsymbol{K})$	Reciprocal space object function for thermally scattered electrons	IV.2.3
$O^{TDS}(\boldsymbol{R})$	real space object function for thermally scattered electrons	IV.2.3
P	degree of polarization	V.3
P	perimeter	VIII.2
$\boldsymbol{P}$	polarization	V.3
$\boldsymbol{P}$	polarization vector	V.2
p	Complement of transmission function	IX.2

Symbol	Designation	Chapter
p	probability of an injected electron reaching the interface	VII.4
p	transition probability	IV.2.5
$P(\boldsymbol{R})$	probe amplitude profile	IV.2.3
$p(s)$	probability distribution of a stochastic quantity s	VIII.1
P_0	polarization of beam of N_0 electrons	V.2
$P^2(\boldsymbol{R})$	probe intensity profile	IV.2.3
$P_{\mathrm{eff}}{}^2(\boldsymbol{R})$	effective probe intensity profile	IV.2.3
P_{fb}	effective spin polarization of the ferromagnetic/barrier interface	VII.2
P_x, P_y	polarization components	V.2
Q	Fourier transform of transmission function	IX.2
Q	ionization cross-section of atoms a	IV.2.5
Q	quality of the cantilever	VII.3
q	electric charge	IV.2.4
q	electron charge density	VIII.1
$\boldsymbol{q}$	momentum transfer on inelastic scattering	IV.2.3
$\boldsymbol{q}$	scattering vector	IX.2
$q(x, y)$	transmission function	IV.2.2
$q(x)$	transmission function	IX.2
$Q_{0(n)}$	energy required to remove surface atom as an ion with charge n in the absence of an electric field	VI.2
$Q_{\mathrm{sp}}^{\mathrm{a}}$, $Q_{\mathrm{st}}^{\mathrm{a}}$	ionization cross-section of element a in specimen, standard	IV.2.5
q_{B}	Brilloin zone boundary	IV.2.3
Q_n	activation energy for surface atom removal as an ion with charge n in the presence of an electric field	VI.2
q_z	momentum transfer parallel to the incident beam direction	IV.2.3
R	effective tip radius	VII.2
R	incident electron beam range	IV.2.1
R	radius	IV.2.5
R	reflectivity of the mirror layer	VIII.1
R	region	VIII.2
R	relative energy resolution; Auger backscattering factor	IV.2.4
$\boldsymbol{R}$	transverse positional coordinate on the specimen entrance surface	IV.2.3
$\boldsymbol{r}$	atomic position	IV.2.3
r	charge separation	VII.3
$\boldsymbol{r}$	incident electron position	V.3
r	radius of curvature; radius of specimen	VI.2
r	tip radius	IX.2
$\bar{r}_0$	center of currature of the effective tip	VII.2
$\boldsymbol{R}_0$	probe position	IV.2.3
r_0	radius of the object field	VIII.1
r_1, r_2	radii of concentric cylinders	IV.2.4
R_{sp}^{a}, R_{st}^{a}	electron backscattering of element a in the specimen, standard	IV.2.5
R_a, R_β	count rate for events a, β	IX.1
$\boldsymbol{r}_{\mathrm{i}}$	target electron position	V.3
$r_{\mathrm{M}}(\mathrm{E}_{\mathrm{A}\alpha})$	backscattering coefficient at angle a to surface normal	IV.2.4
r_{mn}	separation of mth and nth atoms in a column	IV.2.3
S	dissolution rate of the resist development	VIII.1
S	sputter yield	IV.2.6
S	surface	VII.2

Symbol	Designation	Chapter
s	incident electron spin	V.3
s	$\mathbf{K}/4\pi$	IV.2.3
s	normalized arc length	VIII.2
s	single-quantum response	VIII.1
s	tip–sample distance	VII.2
$S(X)$	signal function	IV.2.2
$s^{(k)}$	best search direction	VIII.2
$s(\lambda)$	emission spectrum of a scintillator	VIII.1
$S(\Theta)$	Sherman function	V.2
s(x, y)	three-dimensional tip shape	VIII.2
s, p, d, ...	orbitals within energy levels	IV.2.4
S^a_{sp}, S^a_{st}	electron stopping power of element a in the specimen, standard	IV.2.5
S_A, S_B	area of structures A, B	VIII.2
s_i	target electron spin	V.3
Si(x)	sine integral function	IV.2.3
S_{in}	input signal	VIII.1
$s^j(\boldsymbol{K})$	transverse energy of Block state j	IV.2.3
SNR	signal-to-noise ratio	V.2
S_{out}	output signal	VIII.1
S_{pc}	sensitivity of a photocathode	VIII.1
T	temperature	VI.2, V.1, VII.2, VII.4
T	transmission coefficient of light optics; absolute temperature	VIII.1
T	transmission efficiency to detector in Eq. 1	IX.1
t	flight time	VI.2
t	specimen thickness	IV.2.3, IX.1
t	time	IV.2.6, V.1
$T(u)$	transfer function	IV.2.2
$T(u)$	transfer function of the lens	IV.2.2
$t(x)$, $t(x, y)$	spread function	IV.2.2
$T_{(E_A)}$	transmission efficiency of analyzer at E_A	IV.2.4
$T(E)$	transmission coefficient	VII.2
$t(x)$	spread function	IV.2.2
T_1	spin-lattice relaxation time	V.1
T_2	spin-spin relaxation time	V.1
T_c	critical temperature	IV.2.3
t_d	developing time	VIII.1
$T_{max}(U)$	maximum transmission coefficient	VII.2
T_{zl}	inverse of t_{zl}	IX.2
t_{zl}	parabolic approximation	IX.2
U	accelerating voltage	VIII.1
U	applied sample bias current	VII.2
U	overvoltage	IV.2.5
U	Reciprocal space coordinate	IX.2
u	$(2/\lambda)\sin(\phi)$	IV.2.2
$\boldsymbol{u}$	deviation from atomic position due to thermal motion	IV.2.3
u	spatial frequency	VIII.1
U_0	beam potential	IX.2
V	valence band	IV.2.4

Symbol	Designation	Chapter
V	voltage	VII.4
V	volume	VII.4
v	electron velocity	IV.2.3
v	final velocity of ion	VI.2
V_0'	imaginary part of mean inner potential	IX.2
V_0	threshold voltage	VII.4
$V(\boldsymbol{R})$	specimen projected potential	IV.2.3
$V(r)$	central potential	V.2
V_0	applied voltage	VI.2
$V_1, \ldots V_n$	electrode potentials on a series of electrodes	IV.2.4
V_A, V_B	volume of structures A, B	VIII.2
V_b	ballistic electron emission microscopy voltage	VII.4
V_{ex}	exchange potential	V.3
V_t	injection voltage	VII.4
v_i	potential difference, V_i-V_0	IV.2.4
W_{mn}	correlation factor between mth and nth atoms in a column	IV.2.3
X	a multiplet state resulting from coupling two holes, k and l, in the final state	IV.2.4
$\mathbf{X}$	data matrix	VIII.2
X	distance between incident beam and junction	IV.2.1
X	spatial coordinate on detector	IX.2
x^{-1}, y^{-1}	spatial frequency coordinates	VIII.2
x, y	local window coordinates	VIII.2
x, y	spatial coordinates	VIII.2
x_0	distance from specimen surface	VI.2
X_c	coherence width	IX.2
x_c	critical distance	VI.2
X_n	lateral position of nth fringe	IX.2
$\mathbf{X}^t$	transpose of $\mathbf{X}$	VIII.2
$\mathbf{Y}$	variance-covariance matrix	VIII.2
Z	atomic number	IV.2.1, IV.2.3, IV.2.5, VIII.1
z	cantilever deflection	VII.3
z	coordinate on z axis	IX.2
z	depth	IV.2.5, IV.2.3
z	height of object detail	VIII.2
z	width within which Auger emission occurs	IV.2.4
$Z(x, y)$	atomic number	IV.2.2
z_1	tip-to-sample distance	IX.2
z_2	Distance from sample to detector screen in point projection microscope	IX.2
z_2	sample-to-screen distance	IX.2
α	absorption coefficient of X-rays	VIII.1
α	angle subtended by source at the sample	IX.2
α	angle to surface normal	IV.2.4
α	phase difference between two reflections	IV.2.2
α	weighting coefficient	VIII.2
$a^{\pm}$	transmission efficiency	IV.2.6
a, β	events a, β	IX.1
β	acceptance half-angle of light optics	VIII.1
β	brightness of electron gun	IV.2.1

Symbol	Designation	Chapter
β	coherence angle	IX.2
β	image compression factor	VI.2
β_I	image half-angle of a micro-objective	VIII.1
χ	$(\mu r)\mathrm{cosec}\psi$	IV.2.5
χ	incident electron wavevector	IV.2.3
$\chi\,(u)$	phase factor	IV.2.2
Δ	measure of defocus	IV.2.2
δ	continuum fluorescence contribution/primary characteristic X-ray intensity ratio	IV.2.5
δ	corrugation	VII.2
δ	degeneracy	IX.2
δ	instrumental resolution	IV.2.2
δ	number of secondary electrons per incident electron (-)	IV.2.1
δ	resolution	VI.2
δ	resolution of emission electron microscope	VIII.1
$\delta(\mathbf{R})$	Dirac delta function	IV.2.3
Δ_0	path difference	IX.2
ΔE	change in energy	IV.2.3
ΔE	energy deposit of a primary quantum in a solid	VIII.1
ΔE	energy loss	IV.2.5
ΔE	energy spread	IV.2.4
Δ_E	energy range	IX.2
ΔE_{pe}	energy width of photoelectrons	VIII.1
Δf	objective lens defocus	IV.2.3
δ_{mn}	Kroenecker delta	IV.2.3
$\Delta\mathrm{R}$	transverse separation of two objects	IV.2.3
δ_{s}	secondary emission coefficient	VIII.1
ΔW	Solid angle	IX.2
ε	efficiency of the SEMPA system in Eqs. 6,7	V.2
ε	energy conversion efficiency of a scintillator	VIII.1
$\varepsilon(\omega)$	optical dielectric function	IX.2
ε_0	permittivity of free sface	VI.2
$\varepsilon^j(\boldsymbol{K})$	excitation coefficient of Bloch wave j	IV.2.3
Φ	work function	IV.2.4
ϕ	density of states at the Fermi level for the tip	VII.2
ϕ	scattering angle	IV.2.2
ϕ	Schottky barrier height	VII.4
ϕ	work function	VI.2
$\phi(\Delta rz)$	emission from an isolated thin film of mass thickness Δrz	IV.2.5
Φ_{A}	apparent barrier height	VII.2
Φ_{P}	primary current density	IV.2.6
Φ_{s}	sample work function	VII.2
Φ_{t}	tip work function	VII.2
γ	decay rate	VII.2
γ	fluorescence intensity/primary characteristics X-ray ratio	IV.2.5
γ	quantum yield for impact ionization	VII.4
γ	gyromagnetic ratio	V.1
γ	slope of the characteristic curve of a recorder	VIII.1
$\gamma^{\pm}$	ionization probability	IV.2.6

Symbol	Designation	Chapter
σ	conductance of a tunnel junction	VII.2
σ	cross section of spin-dependent scattering	V.2
σ	electron interaction constant	IV.2.3
σ	interaction constant	IV.2.2
σ	Lennard's constant	IV.2.5
σ	standard deviation	VI.2, VIII.2
$\sigma\uparrow\uparrow$	conductance of a tunnel junction with parallel internal magnetic field directions	VII.2
$\sigma\uparrow\downarrow$	conductance of a tunnel junction with antiparallel internal magnetic field directions	VII.2
$\sigma_{A(Ep)}$	electron impact ionization cross section at primary energy E_p	IV.2.4
σ_a, σ_b	cross section for events a, b	IX.1
σ_{at}	atomic cross-section for scattering to a high angle annular detector	IV.2.3
σ_{fbf}	mean conductance of a tunnel junction	VII.2
σ_t	total inelastic cross-section	IX.2
σ_{true}	cross section for the combined events a, β 2	IX.1
τ	pixel dwell time in Eqs. 6,7	V.2
τ	storage time	VIII.1
τ	time window	IX.1
$\tau^j(\boldsymbol{K}, \boldsymbol{R})$	two-dimensional Bloch state j of the fast electron	IV.2.3
Ω	solid acceptance angle	IV.2.4
Ω	acceptance solid angle of light optics	VIII.1
ω	fluorescence yield; X-ray emission yield	IV.2.5
ω_0	photon energy	V.1
ψ	angle between emitted X-rays and the specimen surface	IV.2.5
ψ $(\boldsymbol{R}, z)$	electron wavefunction within the specimen	IV.2.3
Ψ_h	Electron wavefunction detected in point projection microscope	IX.2
Ψ_i	Electron wavefunction across downstream face of sample	IX.2
$\Psi_i(x)$	electron wavefunction	IX.2
ψ_m	electronic state of the tip	VII.2
Ψ_n	electronic state of the sample	VII.2
ψ_s	amplitude of elastic scattering in the first Born approximation	IV.2.3

Abbreviation	Explanation
3DAP	three-dimensional atom probe
AC	alternating current
ACP	automatic correlation partitioning
ADC	analog-to-digital convertor
ADF	annular dark field
AES	Auger electron spectroscopy
AFM	antiferromagnetism
AP	atom probe
APB	antiphase boundary
BBT	butylbenzotriamide
BF	bright field
BIF	best image field

Abbreviation	Explanation
BSE	backscattered electrons
CCD	charged coupled device
CFD	constant fraction discriminator
CHA	concentric hemispherical analyzer
CID	charge injection device
CL	cathodoluminescence
CLD	constant level discriminator
CMA	cylindrical mirror analyzer
CRT	cathode ray tube
DC	direct current
DD	detector defined
DQE	detective quantum efficiency
EBD	electron beam deposited
EBIC	electron beam induced conductivity
EBIC	electron-beam-induced curent
ECL	emitter-coupled logic
fd	force-distance
FEG	field emission gun
FEM	field emission microscope
FID	free induction decay
FIM	field ion microscope
FM	ferromagnetism
FWHM	full width at half-maximum
GL	gray level
HAADF	high-angle annular dark field
HDTV	high-definition television
HOLZ	High order Laue zone
HREM	high resolution electron microscopy
ICP	interactive correlation partitioning
ISIT	intensified silicon intensifier target
LDOS	local density of states
LEDS	low energy diffuse scattering
LEED	low-energy electron diffraction
LIMA	laser ionization mass analysis
LLL	low light level
MCA	multichannel analyzer
MCD	magnetic circular dichroism
MCP	microchannel plate
MCS	multichannel scaler
MIS	metal/insulator/semiconductor
MISR	matrix/ion species ratio
MLCFA	maximum likelihood common factor analysis
MMF	minimum mass fraction
MOSFET	metal oxide semiconductor field effect transistor
MSA	multivariate statistical analysis
PC	personal computer
PHD	pulse height distribution
PLAP	pulsed laser atom probe
PMMA	polymethylmethacrylate

Abbreviation	Explanation
PMT	photomultiplier tube
PoSAP	position-sensitive atom probe
PSF	point spread function
QSE	quantum size effect
QUANTITEM	quantitative analysis of the information from electron micrographs
RAE	resistive anode encoder
RF	radio-frequency
SCA	single channel analyzer
SD	source defined
SE	secondary electrons
SEC	secondary electron conduction
SED	secondary electron detectors
SIT	silicon intensifier target
SKIZ	skeleton by influence zones
SNR	signal-to-noise ratio
SSCCD	slow-scan charge-coupled device
TAC	time-to-amplitude converter
TAP	topographic atom probe
TOF	time of flight
TTL	transistor-transistor logic
TV	television
UHV	ultrahigh vacuum
VCR	video cassette recorder
YAG	yttrium aluminum garnet

Techniques	Explanation
AEEM	Auger electron emission microscopy
AEM	analytical electron microscopy
AES	atomic emission spectroscopy
AES	Auger electron spectroscopy
AFM	atomic force microscopy
APFIM	atom probe field ion microscopy
ARPES	angle-resolved photoelectron spectroscopy
ATRS	attenuated total reflectance spectroscopy
BEEM	ballistic electron emission microscopy
BEES	ballistic electron emission spectroscopy
BF CTEM	bright field conventional transmission electron microscopy
CBED	convergent beam electron diffraction
CBRHEED	continuous beam reflective high-energy electron diffraction
CCSEM	computer-controlled scanning electron microscopy
CITS	current imaging tunneling spectroscopy
CL	cathodoluminescence
CLSM	confocal laser scanning microscopy
CT	computer-aided tomography
CTEM	conventional transmission electron microscopy
CVD	chemical vapor deposition

Techniques	Explanation
DLTS	deep level transient spectroscopy
EBT	electron beam testing
EDS	electron diffraction spectrometry
EDS	energy-dispersive spectroscopy
EDX	energy dispersive X-ray spectroscopy
EELS	electron energy loss spectroscopy
EFTEM	energy filtered transmission electron microscopy
EM	electron microscopy
EPMA	electron probe microanalysis
EPXMA	electron probe X-ray microanalysis
ESCA	electron spectroscopy for chemical analysis
ESEM	environmental scanning electron microscopy
ESI	electron spectroscopic imaging
ESI	element-specific imaging
FFM	friction force microscopy
FIB	focused ion beam milling
FIM	field ion microscopy
FMT	fluorescent microthermography
FT-IR	Fourier transform infrared spectroscopy
HREM	high resolution electron microscopy
HRSEM	high resolution scanning electron microscopy
HRTEM	high resolution transmission electron microscopy
HVEM	high voltage electron microscopy
LACBED	large angle convergent beam electron diffraction
LCT	liquid crystal thermography
LEEM	low-energy electron microscopy
LFM	lateral force microscopy
LM	light microscopy
LMMS	laser microprobe mass spectrometry
LOM	light optical microscopy
LPCVD	low-pressure chemical vapor deposition
LTSLEM	low-temperature scanning laser electron microscopy
M-PIXE	micro-(proton-induced X-ray emission spectroscopy)
MBE	molecular beam epitaxy
MEM	mirror electron microscopy
MFM	magnetic force microscopy
MOVPE	metal-organic vapor phase epitaxy
MRI	magnetic resonance imaging
MULSAM	multispectral Auger microscopy
NMR	nuclear magnetic resonance
OM	optical microscopy
PCA	principal components analysis
PEELS	photoelectron energy loss spectroscopy
PEEM	photoemission electron microscopy
PFA	principal factor analysis
PIXE	proton induced X-ray emission spectroscopy
PL	photoluminescence
PPM	point-projection microscopy
RBS	Rutherford backscattering spectroscopy

Techniques	Explanation
RDE	reactive deposition epitaxy
REM	reflection energy microscopy
REM	reflection electron microscopy
RHEED	reflection high-energy electron diffraction
SAM	scanning acoustic microscopy
SAM	scanning Auger microscopy
SAXS	small-angle X-ray scattering
SCM	scanning capacitance microscopy
SDLTS	scanning deep level tansient spectroscopy
SECM	scanning electrochemical microscopy
SEELS	serial electron energy-loss spectroscopy
SEEM	secondary electron emission spectroscopy
SEM	scanning electron microscopy
SEMPA	scanning electron microscopy with polarization analysis
SFM	scanning force microscopy
SIMS	secondary ion mass spectrometry
SLEEM	slow low-energy electron microscopy
SNOM	scanning near-field optical microscopy
SNPM	scanning near-field probe microscopy
SPE	solid phase epitaxy
SPLEED	spin-polarized low-energy electron diffraction
SPLEEM	spin-polarized low energy electron microscopy
SPM	scanning probe microscopy
SPSTM	spin-polarized scanning tunneling microscopy
SQUID	superconducting quantum interference device
SREM	scanning reflection electron microscopy
STEM	scanning transmission electron microscopy
STM	scanning tunneling microscopy
STOM	scanning tunneling optical microscopy
STS	scanning tunneling spectroscopy
STXM	scanning transmission X-ray microscopy
TED	transmission electron diffraction
TEEM	thermionic electron emission microscopy
TEM	transmission electron microscopy
TL	thermoluminescence
TS	tunneling spectroscopy
TSMFM	tunneling stabilized magnetic force microscopy
TXM	transmission X-ray microscopy
UFM	ultrasonic force microscopy
UMT	ultra microtomography
VPE	vapor phase epitaxy
WDS	wavelength dispersive spectroscopy
XES	X-ray emission spectroscopy
XPS	X-ray photoelectron spectroscopy
XPS	X-ray photoemission spectroscopy
XTEM	cross-sectional transmission electron microscopy

Handbook of Microscopy
Edited by S. Amelinckx, D.van Dyck, J.van Landuyt ,G.van Tendeloo

Index